THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 54, PART 3 1 May 1942 No. 303

WILLIAM DAVID COOLIDGE

EIGHTEENTH DUDDELL MEDALLIST

THE eighteenth Duddell Medal was presented to Dr. W. D. Coolidge by His Excellency the British Ambassador, the Right Honourable Viscount Halifax, at a dinner given on 1 May during the meeting of the American Physical Society at Baltimore.

Coolidge is best known to physicists for his x-ray tube, and for his share in the development of the tungsten filament, either of which falls within the category of "development of instruments or materials", which it is the prime purpose of the Duddell Medal to recognize.

William David Coolidge, this year's medallist, was born in 1873 and passed his early life on a farm in Massachusetts; he graduated at the Massachusetts Institute of Technology in 1896, at a time when the first announcement of the discovery of x rays was exciting interest everywhere. He remained at the M.I.T., except for a short period of research at Leipzig, until 1905, when he went to the research laboratories of the General Electric Company, Schenectady, where he has remained ever since. He became Assistant Director of the Laboratories in 1908, Associate Director in 1932, and finally a vice-president of the company and Director of Research since 1940.

His work on tungsten, as is well known, led to the possibility of producing this usually brittle metal in ductile form, and so made it possible to use it in electric lamps, where its high melting point permits the filament to run at a very high temperature, so as to obtain great luminous efficiency. The value of the work to physicists, however, has been not only in the possibilities which it opened up in the field of instrument construction, but because it focused attention on some of the problems of the solid state. In this way, his work has contributed indirectly to this fundamental problem of theoretical physics.

His hot-cathode x-ray tube was described 30 years ago; it differed from earlier tubes in that the current and voltage could be controlled independently. In his tube, which is very highly evacuated, variation of the current through the filament alters the electron emission from the cathode, and thus gives control of the tube current, and with it of the intensity of the radiation. The voltage between anode and cathode can be varied independently of this, thus enabling radiation of varying "hardness" to be produced. Later, Coolidge developed oil-immersed self-contained x-ray units, which have been of great value in medical and industrial work. A further development was announced in 1924, when he described a tube for high voltages. This tube was sub-divided, and accelerating voltages applied to the various sections. In the accompanying

photograph, the designer is seen holding a tube designed for one million volts. Such tubes have enabled X rays of great penetrating power to be applied medically, and have enabled deep therapy to be used under closely controlled conditions. They have also been applied industrially in radiographic examination of materials.

Another outcome of Coolidge's work has been the study of the cathode rays outside the tube, a development which aroused considerable interest both in the

Coolidge inspecting his million-volt x-ray tube.

technical and the lay press a few years ago. The cathode-ray particles emerge from the tube through an exceedingly thin window of metal foil, and their track in air is brilliantly marked. In living tissue, their penetration with the voltages available at first was very small, but with higher exciting voltages, rays of great penetrating power may be anticipated. These promise to be of therapeutic value in certain maladies.

THE PHYSICS OF BLOOD-PRESSURE MEASUREMENT

By DAVID S. EVANS * and K. MENDELSSOHN,
Clarendon Laboratory, Oxford

MS. received 17 *December* 1941

ABSTRACT. The physical basis of blood-pressure measurement has been formulated. An arrangement for recording the essential features of the pulse curve is described, and the relation between the recorded values and the pressure changes in the artery discussed. The results have been confirmed by clinical observations.

§ 1. INTRODUCTION

In wartime, the determination of arterial blood-pressure in casualties is a matter of the greatest importance, since it is the only means for the early detection of the serious condition known as shock. It seems, therefore, worth while to discuss in detail the physical basis of such determinations and to explore the possibilities of the development of existing methods in order to improve their accuracy and reproducibility.

The action of the heart produces in the arteries a pressure variation which is more or less periodic. Medical practice is accustomed to define three parameters in connection with this variation. They are :—

(i) The frequency, or *pulse rate.*

(ii) The maximum pressure attained in the artery during the pulsation. This is the *systolic pressure*, p_s.

(iii) The minimum, or *diastolic pressure*, p_d.

The quantity, $p_s - p_d$, is called the *pulse pressure*, which we have denoted by p_0.

Although the description of the pulse curve given by these parameters is rather rough, they are the only quantities which have hitherto been readily determinable in practice. Typical values for a healthy individual are:— Systolic pressure: 100 to 150 mm. of mercury, with pulse pressure about one-third of systolic. Pulse rate: 60 to 120 per minute. In cases of shock the systolic pressure may be as low as 40 to 60 mm.

For practical reasons it is impossible to tap the artery directly in order to measure the pressure. The usual arrangement for measurement consists of a rubber cuff, A (see figure 1), which can be inflated with air to any desired pressure. It is secured in position around the limb l by means of an inextensible bandage. The air pressure in the cuff is read on the manometer m_1. For our purposes, the tissues, t, of the limb may be regarded as an incompressible jelly, which

* By kind permission of the Board of Visitors of the University Observatory, Oxford.

transmits the external pressure applied by the cuff. Thus, when the pressure in A is raised to a value equal to the systolic pressure in the artery a, the latter collapses completely throughout the pulse cycle, and cuts off the flow of blood. The systolic pressure can, therefore, be determined with the aid of any device which will indicate the cessation of blood-flow past A, the pressure at which this happens being read on m_1.

The usual method is to employ a stethoscope for detection, and systolic pressure is indicated by the first appearance of pulse sounds as the external pressure is lowered. The diastolic pressure is read in a similar manner, the indication in this case being that the arterial sounds grow muffled or disappear entirely when the pressure in A falls to a certain value.

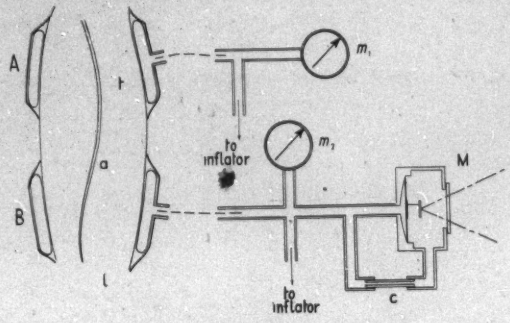


Figure 1.

The reason for this change in the sounds is unknown, but we think it may be due to a change from turbulent to laminar flow in the artery. There exists no theoretical reason for identifying the pressure at which this change in sound occurs with the diastolic pressure as defined. Besides being theoretically unsound, the auscultatory method has the disadvantage of yielding values on the same patient which vary widely with individual measuring technique, and it is difficult to carry out in an operating theatre or casualty clearing station, where conditions are not conducive to the accurate detection of faint sounds in a stethoscope.

Amplification of the pulse sounds, or their conversion into a visual record by electrical methods, is ruled out, because simplicity and cheapness are essential factors in the solution of the practical problem of producing a recording instrument.

§ 2. EXPERIMENTAL ARRANGEMENT

The detector chosen after a number of exploratory trials was a second cuff, B, which was placed on the limb below A.*

* The use of a second cuff, or the combination of two cuffs, is by no means new, but the various modes of operation hitherto employed are fundamentally different from ours.

This recording cuff, B (figure 1), is exactly similar to A, and consists of a rubber bag and an inextensible bandage to strap it around the limb. It can be inflated to any desired pressure, which is indicated on the manometer m_2. Pulsations in the artery are transmitted to this cuff, and are measured by means of a highly sensitive membrane manometer, M, whose deflections are recorded by means of a light beam and scale. This manometer consists of two chambers, of which the smaller is directly connected to the cuff, while the connection with the larger is made through a capillary, c. This arrangement serves to record only the pressure changes of short period. If the dimensions of the capillary and manometer are correctly chosen (see § 4), it records the patient's pulse faithfully, but is undisturbed by the very much larger, but gradual, volume changes due to the variation of temperature or fluid content of the tissues.

§ 3. THE DETERMINATION OF p_s, p_d AND p_0

We shall now discuss in detail the relation between the record provided by the instrument and the true values of systolic, diastolic and pulse pressure in the artery.

We consider first the amplitude of the pressure pulsations in the inflated cuff B, as indicated by M, in relation to the mean pressure in B. The upper cuff, A, is left deflated. We define the following quantities:—

p, the pressure within the artery at any instant.
P, the mean pressure of air in the cuff B.
v, the volume at any instant of that part of the artery covered by B.
V, the total volume of B and the recording circuit.
M, the mass of air contained by V.

Since the cuff is restrained by an inextensible bandage, any increase in the volume of the artery must be compensated by an equal decrease in the volume of the cuff, or

$$v + V = \text{constant}. \qquad \ldots\ldots(1)$$

Further,

$$PV = kM, \qquad \ldots\ldots(2)$$

where k is a constant for the system under consideration. If, during the pulsation, the volume of the artery increases by an amount Δv, there will be a corresponding increase of pressure ΔP in the cuff, given by

$$\Delta P = P.\Delta v/V, \qquad \ldots\ldots(3)$$

because Δv is very small in comparison with V.

The artery is an elastic tube of negligible rigidity, so that the arterial volume v is given by

$$\left.\begin{aligned} v &= v_0 + \lambda(p - P) \text{ when } p > P \\ &= 0 \qquad\qquad\quad \text{when } p < P, \end{aligned}\right\} \qquad \ldots\ldots(4)$$

where v_0 and λ are constants depending on the individual patient.

The changes in arterial volume caused by the pulse are of three different types corresponding to the three conditions:—

(i) External pressure less than the diastolic pressure ($P<p_d$),

(ii) External pressure intermediate between systolic and diastolic pressures ($p_d<P<p_s$),

(iii) External pressure greater than systolic ($p_s<P$).

These three cases correspond respectively to the regions I, II and III in figure 2, which illustrates the relation of the amplitude, ΔP to P, the mean pressure in the cuff B.

In case (i), the artery merely pulsates freely between the extreme values $v_0+\lambda(p_s-P)$ and $v_0+\lambda(p_d-P)$ given by equation (4), so that

$$\Delta v=\lambda(p_s-p_d)=\lambda p_0, \qquad \ldots\ldots(5)$$

and hence in this case

$$\Delta P=\frac{P}{V}.\lambda p_0, \qquad \ldots\ldots(6)$$

i.e., the amplitude of oscillation increases linearly with increasing external pressure, and tends to the value $\lambda p_0 p_d/V$ as P approaches diastolic pressure. This is represented by the straight line shown in region I of the figure.

In case (ii) the change in volume of the artery during the pulse cycle is from $v_0+\lambda(p_s-P)$ at systolic pressure to zero volume at diastolic, so that now

$$\Delta P=\frac{P}{V}(v_0+\lambda p_s-\lambda P). \qquad \ldots\ldots(7)$$

This relation represents the amplitude of the pressure oscillations in the cuff between the values $P=p_d$ and $P=p_s$, subject to certain reservations which will presently be discussed.

The curve of ΔP against P in region II is a parabola, which attains its maximum at

$$P=\frac{v_0}{2\lambda}+\frac{p_s}{2}, \qquad \ldots\ldots(8)$$

which is greater than p_s if $v_0>\lambda p_s$, and this appears to be the case, since on the whole the variations of arterial volume in an artery not subject to external interference are rather small compared with its total volume. Equation (7) shows that as P approaches p_d from above, ΔP tends to the value $\frac{p_d}{V}(v_0+\lambda p_0)$, while we have seen that if diastolic pressure is approached from below, ΔP attains the value $\lambda p_0 p_d/V$. Thus the amplitude ΔP of the pressure oscillations in B passes through a discontinuity at a value of P equal to the true diastolic pressure in the artery. The change in amplitude at this discontinuity is given by $v_0 p_d/V$. The occurrence of this discontinuity provides us with an ideal method for the determination of accurate values of the diastolic pressure, and has been successfully employed for this purpose in our clinical work.

We now consider case (iii). It might be expected that there would be no pulsations in B when the external pressure was greater than systolic, since the

artery should then be completely collapsed throughout the pulse cycle. This, however, is not the case, since the pressure exerted by the cuff decreases in the neighbourhood of its edges. It follows that the short section of artery covered by the upper edge of B fills with blood at the peak of each pulse wave, and causes pressure oscillations in B even when $P>p_s$. The amplitude does not therefore fall abruptly to zero as P crosses p_s, but dies away gradually.

A second disturbing effect is of importance in the neighbourhood of systolic pressure. The artery below B remains empty during the greater part of the pulse cycle, and the part of the cycle during which the artery is open is so short that the flow of blood at each pulse beat is insufficient to fill completely the section of artery covered by B.

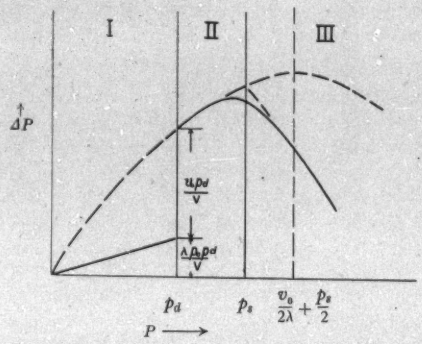


Figure 2.

The amplitude, therefore, decreases below the value given by the parabola even before p_s is reached (see figure 2). The situation is not well adapted to mathematical description, and p_s cannot be determined with any accuracy using only a single cuff.

For the determination of the systolic pressure we have therefore employed the cuff A as well as B. The latter is inflated to a moderate pressure below diastolic, and then serves as a detector for blood flowing past A. If the pressure in A is allowed to fall through the systolic value, pulsations will start suddenly in B as soon as blood begins to flow past A, thus allowing systolic pressure to be determined. With this arrangement edge effects are confined to A, and the determination is a critical one.* By observing the sudden changes in the recorded pulsations in relation to the pressures in cuffs A and B we are, therefore, able accurately and unambiguously to determine systolic and diastolic pressure, the two parameters on which current medical practice relies. It appears, however,

* Cuff A must, of course, be wide enough to exert an even pressure over a considerable length of the artery.

from an analysis of the (ΔP, P) curve that our arrangement will be capable of recording further valuable information concerning the patient's circulation. Reference to equation (6) shows that when the external pressure is less than p_d the amplitude of oscillation in B is proportional to p_0. Now p_0 is the pulse pressure, which gives a direct indication of the quantity of blood flowing through the limb at each heart beat. Medical opinion attaches an increasing importance to this quantity, since a drop in pulse pressure seems to be the earliest and most reliable indication of the onset of shock. Previously p_0 could only be deduced from separate determinations of p_s and p_d, a process involving a double error and a considerable expenditure of time and attention. Our arrangement enables the observer to form a continuous estimate of p_0 itself at each moment, from the amplitude of the oscillations recorded by the manometer when B is inflated to a constant pressure below diastolic.

§ 4. THE RECORDING OF THE PULSE CURVE

So far we have been concerned only with the amplitude of the pressure oscillations in B. We now consider their detailed form as well. We take the case when the external pressure is less than p_d, and consider an instant when the arterial pressure is p. Then the arterial volume is given by $v_0+\lambda(p-P)$; at diastolic pressure it is given by $v_0+\lambda(p_d-P)$, so that the change of volume is $\lambda(p-p_d)$, and hence the pressure change in the cuff is, by equation (3),

$$\Delta P=\frac{P}{V}.\lambda(p-p_d). \qquad \ldots\ldots(9)$$

Thus, as the arterial pressure p alters with time, the pressure variations in the cuff trace out a curve which is similar to the variation of pressure within the artery. This relation at once suggests that it should be possible to produce an instrument to record accurately all the pressure variations within the artery throughout the pulse cycle, and immediately raises the problem of the faithfulness of the record provided by our apparatus. Distortion of this record arises both from the action of the capillary and from the natural period and damping of the moving parts of the manometer.

The volumes of the two chambers of the manometer M are denoted by V_1 and V_2, and the pressures in them are P_1 and P_2. The division between the two chambers (see figure 1) is a metal membrane, and they are connected by a capillary of length l and diameter d, through which excesses of pressure in the one chamber leak slowly into the other, until a mean pressure P_0 is attained. The difference P_1-P_2 is very small, and never exceeds one or two mm. of water.

Then we have

$$P_1V_1+P_2V_2=P_0(V_1+V_2). \qquad \ldots\ldots(10)$$

If P_1 is greater than P_2, air will flow from one chamber to the other at a rate given by $\kappa(P_1-P_2)$ c.c. per second (assuming laminar flow), where

$$\kappa=\pi d^4/128\eta l \qquad \ldots\ldots(11)$$

and η is the viscosity of air ($=1{\cdot}8\times10^{-4}$ c.g.s. units). This volume is measured at the mean pressure, which we may take as P_0, since P_1-P_2 is very small. Then we have the equations

$$\frac{dP_1}{P_0}=-\frac{\kappa}{V_1}(P_1-P_2)dt;\quad \frac{dP_2}{P_0}=\frac{\kappa}{V_2}(P_1-P_2)dt, \qquad \ldots\ldots(12)$$

of which equation (10) is an integral.

Elimination of P_2 from the first of (12) gives

$$dP_1=-\frac{2\kappa}{V_0}(P_1-P_0)dt, \qquad \ldots\ldots(13)$$

where

$$\frac{2}{V_0}=\frac{1}{V_1}+\frac{1}{V_2},$$

and so

$$\left.\begin{aligned} P_1-P_0&=(P_1{}^0-P_0)e^{-\phi t},\\ P_2-P_0&=(P_2{}^0-P_0)e^{-\phi t},\\ \phi&=2\kappa P_0/V_0, \end{aligned}\right\} \qquad \ldots\ldots(14)$$

where $P_1{}^0$ and $P_2{}^0$ are the initial pressures.

If $F(t)$ is the pressure variation with time recorded by the manometer, and $f(t)$ the true pressure variation in the cuff, then

$$\frac{dF(t)}{dt}=\frac{df(t)}{dt}-\phi F(t), \qquad \ldots\ldots(15)$$

since the pressure decays at a rate given by $-\phi F(t)$.

Integrating (15) from $-\infty$ to t gives

$$\begin{aligned} F(t)&=e^{-\phi t}\int_{-\infty}^{t}e^{\phi t}\frac{df(t)}{dt}\,dt\\ &=f(t)-\phi\int_0^{\infty}f(t-x)e^{-\phi x}dx, \qquad \ldots\ldots(16) \end{aligned}$$

where x is a dummy variable.

When ϕ is small (i.e. the capillary is very narrow), $F(t)$ approximates to $f(t)$, but we cannot dispense with the capillary entirely, since our highly sensitive manometer must act differentially if it is to record small fluctuations superposed on a very high mean pressure liable to slow changes. The practical limit to the smallness of ϕ is set by the necessity of preventing serious deviations of mean pressure between the two chambers. As ϕ increases, $F(t)$ tends to zero, and since ϕ is proportional to P_0, the operating pressure should be as low as possible for accurate reproduction of the pulse curve. On the other hand, equation (9) shows that the sensitivity is proportional to the cuff pressure, so that ϕ should be as small as possible, and the apparatus should be operated as near as possible to diastolic pressure.

If $V_1=V_2=10$ c.c., and $P_0=30$ mm. of mercury, then a capillary 5 cm. in length gives $\phi=21{\cdot}5$ when $d=1$ mm., $\phi=1{\cdot}34$ when $d=\frac{1}{2}$ mm. and $\phi=0{\cdot}27$ when $d=\frac{1}{3}$ mm.

In figure 3, the effect on a typical pulse-pressure curve is illustrated. The pulse rate is taken to be 1 per second, so that if the pulse rate is 120 per minute, the values of ϕ given above must be halved, i.e., more rapid pulses are recorded more accurately. We see that $\phi=1$ already gives quite a good reproduction of the curve form. If we are merely interested in the amplitude, as we are when measuring p_s, p_d or p_0, a much wider capillary can be used, since the reduction in amplitude from $\phi=1$ to $\phi=10$ is not serious. The recorded amplitude is proportional to the true amplitude, so that the analysis of the $(\Delta P, P)$ curve made in §3 still holds.

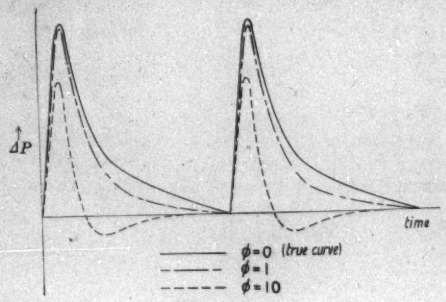


Figure 3.

Calculation of the distortion produced by the natural period and damping of the instrument shows that good reproduction can be obtained with a period equal to 1 % of the pulse period, but that a ratio of only 10 to 1 may be tolerated if the damping is considerable.

§5. CONCLUSION

An experimental model embodying the principles discussed in this paper has been constructed and used in extensve clinical tests. A detailed description of the apparatus and the clinical results will appear in a medical journal. These results have borne out the features of the theoretical $(\Delta P, P)$ curve (figure 2). The continuous observation of p_0 has been carried out during lengthy operations, and has proved of considerable value both to anaesthetist and surgeon. No use has been made so far of the detailed form of the pulse curve whose observation was discussed in §4, since it was our aim to produce a recording apparatus simple and cheap enough to permit mass production. There is, however, no doubt that an accurate knowledge of the pulse curve will be of great importance in many cases, and this record could easily be obtained from our arrangement. The production of a second model fitted with a film camera is under consideration.

§6. ACKNOWLEDGEMENTS

In conclusion, we wish to express our thanks to Professor R. R. Macintosh, of the Department of Anaesthetics, Oxford University, who formulated the medical problem, and to the staff of his department for their interest and help in the clinical experiments. We are also indebted to the surgeons of the Radcliffe Infirmary, Oxford, who accorded us facilities in the operating theatre. Finally, our thanks are due to Mrs. V. Launchbury, whose voluntary assistance has been invaluable throughout the experimental work.

COLOUR BLINDNESS AND ITS IMPORTANCE IN RELATION TO INDUSTRY*

By F. H. G. PITT,
Developments Dept., Kodak Ltd.

MS. received 29 *January* 1942

It is common knowledge to all persons interested in colour that the colour-vision characteristics of individuals vary considerably. This variation is so great that whereas one observer will say that two colours are completely different, another may say that they are almost identical. Persons whose colour vision is different from that of the normal, and whose range of colours is in general more limited, may be broadly defined as being colour blind. While this term is rather misleading, the implication being that all colour-blind persons are blind to colour, it has been so long and so universally used in the literature on the subject as to render its replacement by a better term almost impossible. Colour blindness is almost always congenital, and as such is incurable. There is good evidence that colour blindness may be acquired, but reliable information on this aspect of the subject is very scarce, and for this reason only the inherited type will be discussed.

Much information has been collected on normal and abnormal colour-vision, and it is possible, without any reference to the numerous colour-vision theories which exist, to classify scientifically the types of colour vision of the human race into three main classes, which may be in turn further subdivided. This classification may be arranged as follows:—

Trichromats

This class contains those whose colour vision is a function of three variables, it being possible for an observer in this class to express any colour, C, uniquely by an equation of the type

$$C = \alpha R + \beta G + \gamma B,$$

* A paper read to the Colour Group 11 December 1941

where R, G and B may, for example, be the primaries of a colorimeter and α, β and γ the amounts required to match the colour C.

This class may be subdivided into

(*a*) Persons with normal colour-vision,

(*b*) Persons with abnormal colour-vision, called Anomalous Trichromats.

Dichromats

This class contains those whose colour vision is a function of two variables, and may be subdivided into

(*a*) Protanopes, called red-blind by Helmholtz and red-green blind by Hering;

(*b*) Deuteranopes, called green-blind by Helmholtz and red-green blind by Hering;

(*c*) Tritanopes, called blue-blind by Helmholtz and blue-yellow blind by Hering.

Von Kries introduced the Greek names, which were meant to be descriptive only.

Monochromats

This class contains those whose colour vision is a function of one variable only; cases are extremely rare. This class, together with that of Tritanopes, also rare, are not deemed numerous enough to be taken into account in a paper of this description. As both classes are generally due to disease of the eye, their defect should never go undetected.

Colour blindness is only abnormal when referred to normal vision. It is, therefore, necessary to compare the results from both types for a full understanding of the subject. As it will be shown later that anomalous trichromatism may yield varying characteristics, whereas those of the dichromats form a fairly well-defined set, comparison between the latter class and normal vision will be made first. It is obvious that this comparison is best obtained by the knowledge of how these classes of observers see certain colours. The spectrum, with its large range of pure colours of varying intensities, which can be produced under closely controlled conditions, forms an admirable test object. It is possible for the observers to measure the change of hue, the intensity, and the relative amounts of certain pre-determined primaries from one wave-length to another and thus record exactly how the spectrum appears to them. This information is of extreme importance, for on it is based—knowingly, or sometimes unknowingly—all tests used in the detection of colour blindness. It has the advantage of paying no attention to nomenclature, which undoubtedly differs markedly from observer to observer.

Hue discrimination

The measurement of hue-discrimination data offers more difficulty than that of luminosity or colour mixture. The proper conditions of observations

are important, and it is essential that the instrument used should be easily controlled by the observers, as this type of measurement is of a tiring kind. Unless these conditions are observed, and unless specially suitable apparatus is used, the results tend to become a function of variables other than the actual observers. In practice, the method used is to find the just-perceptible difference between two juxtaposed fields in terms of wave-length. Results shown in

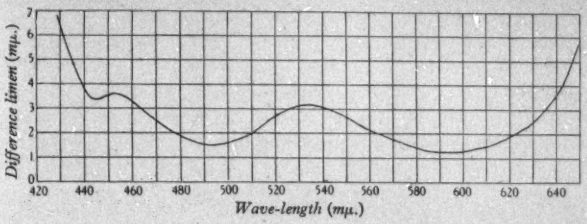


Figure 1. Hue-discrimination curve for normal observer (Wright and Pitt, 1934).

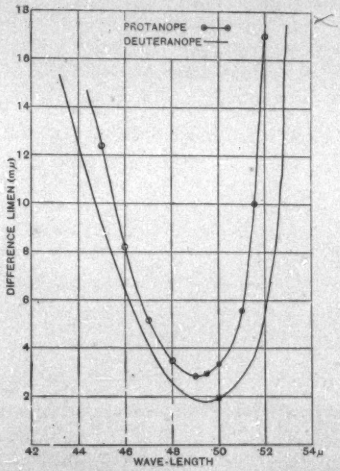


Figure 2. Mean hue-discrimination curves from the results of 6 protanopes and 6 deuteranopes (mean position of neutral points marked with asterisks) (Pitt, 1935).

figures 1 and 2 are for normal trichromatic and dichromatic observers, for a 2° field, and were all made by means of the same apparatus, the Wright Colorimeter (Wright, 1927/28), on which, incidentally, all the other measurements described were made.

The differences between the trichromatic and dichromatic observers are

probably the most outstanding of all the differences. The curves are in themselves self-explanatory, and from them it can be calculated that the average deuteranope can see about 30 just-distinguishable spectral colours and the average protanope about 20 just-distinguishable spectral colours, while the normal trichromat can distinguish about 150 spectral colours. This comparison between trichromat and dichromat does not, however, present the full story. Colour for dichromats is a function of two variables only. As one coefficient increases the other correspondingly decreases, and it therefore follows that all colours possible to dichromatic vision, whether spectral or otherwise to the dichromats, can be seen by dichromats in the spectrum. It has been estimated, as a rough approximation, that the normal trichromat can distinguish about 5000 different colours when a 2° matching field is used. In comparison, the dichromat lives in a very limited world of colour (see Konig, 1903; Steinder, 1906; Jones, 1917; Laurens and Hamilton, 1923; Wright and Pitt, 1934; Hecht and Schlaer, 1936).

Luminosity

The standard luminosity or visibility curve for an equal-energy spectrum, as proposed by Gibson and Tyndall (1923), was adopted by international agreement at the meeting of the International Commission of Illumination of 1924. This curve does not represent the observations of one particular person, but

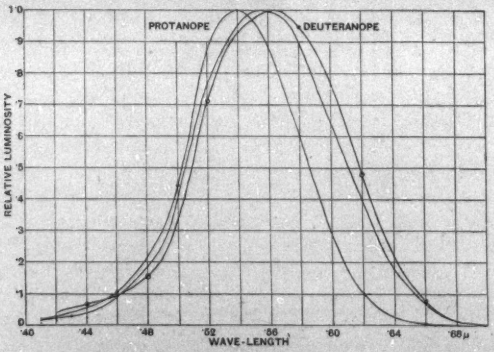


Figure 3. Mean luminosity curves (equal-energy spectrum) from the results of 6 protanopes and 6 deuteranopes; also normal curve (Pitt, 1935).

was deduced from the data of various researches on the subject, and really represents the average eye of 250 normal trichromatic observers. In figure 3 a normal curve is shown together with the mean luminosity curves from the results of 6 protanopes and 6 deuteranopes.

The essential differences of the dichromatic curves are apparent. The protanopic curve has a sharper maximum, the shape of the curve being narrower

throughout. Probably the most important point to notice is that the ordinates of the protanopic curve are considerably smaller from 0·56 μ to the red end of the spectrum: this is the only essential difference between the protanope and deuteranopes. From 0·53 μ to the blue end of the spectrum the differences between all three luminosity curves are not marked, and the curves undoubtedly overlap for individuals in the three groups. The red ordinates of the deuteranope tend to be higher than those of the normal trichromat (see Pitt, 1935; Wright and Pitt, 1935; Hecht and Schlaer, 1936).

Colour mixture

For the normal observer, any colour may be represented in terms of three other colours. That is,

$$C = \alpha R + \beta G + \gamma B,$$

where C is any colour, and R, G and B are three other colours. In modern colorimetric work the sum of the coefficients α, β and γ is made to equal unity, as this specifies the quality of the colour C uniquely, although no indication of the luminosity is given.

In the same way, for the dichromatic observer, any colour may be represented in terms of two other colours. That is,

$$C = \alpha R + \beta B.$$

The data of figures 4 and 5 used the W.D.W. system of units (Wright, 1929), and can be transformed readily to the better known C.I.E. system (Guild, 1924/25;

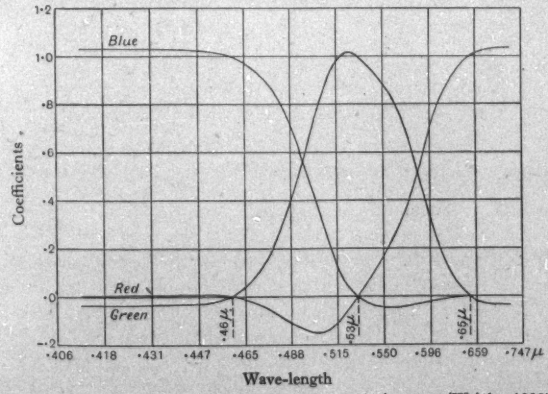


Figure 4. Mean coefficient curves from results of 10 observers (Wright, 1929).

Pitt, 1935). Figure 4 is the mean curve calculated from the results of 10 normal observers. Mean curves from the results of 8 protanopes and 7 deuteranopes are shown in figure 5. The differences between figures 4 and 5 are to be

expected from the foregoing remarks. The two mean curves shown in figure 5 show a difference from $\lambda = 0{\cdot}51\mu$ to the red end of the spectrum, but it should be pointed out that the coefficient curves cannot be used as an infallible test to show

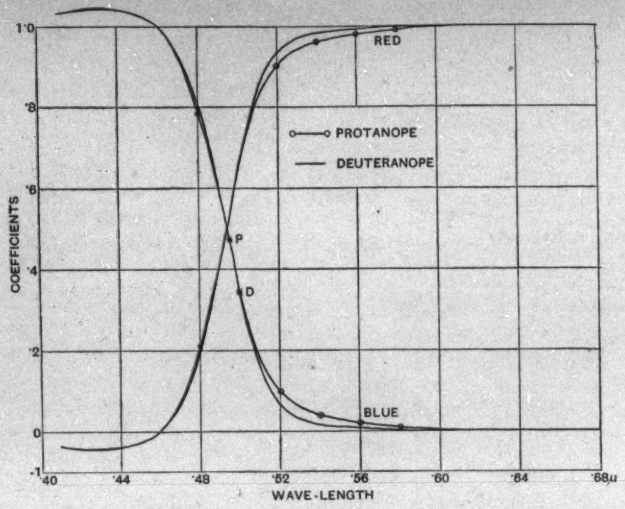


Figure 5. Mean coefficient curves from the results of 8 protanopes and 7 deuteranopes (position of neutral point marked with asterisks) (Pitt, 1935).

whether the dichromat in question is a protanope, as the individual curves of both classes overlap (see Konig and Dieterici, 1892; Abney, 1913; Wright, 1929; Pitt, 1935.)

Anomalous trichromats

In general, anomalous trichromats differ from normal trichromats in that, although they have a colour system which is a function of three variables, the colour matches may differ from those of a normal trichromat. It is not possible to present the results of measurements made by anomalous trichromats in the same way as those of the normal trichromats or dichromats have been presented, as anomalous trichromatism ranges progressively between normal trichromatism and dichromatism (Nelson, 1938; McKeon and Wright, 1940). Figure 6, due to McKeon and Wright, shows hue-discrimination curves of 11 protanomalous observers arranged in order of increasing magnitude of defect. From these curves it can be seen that hue discrimination becomes progressively poorer in the greenish-yellow part of the spectrum. McKeon and Wright also state that the luminosity data vary little from observer to observer, and the mean curve is

very similar to that of the protanope. In figure 7, due to Nelson, is shown hue discrimination of 5 deuteranomalous observers. Once again the general

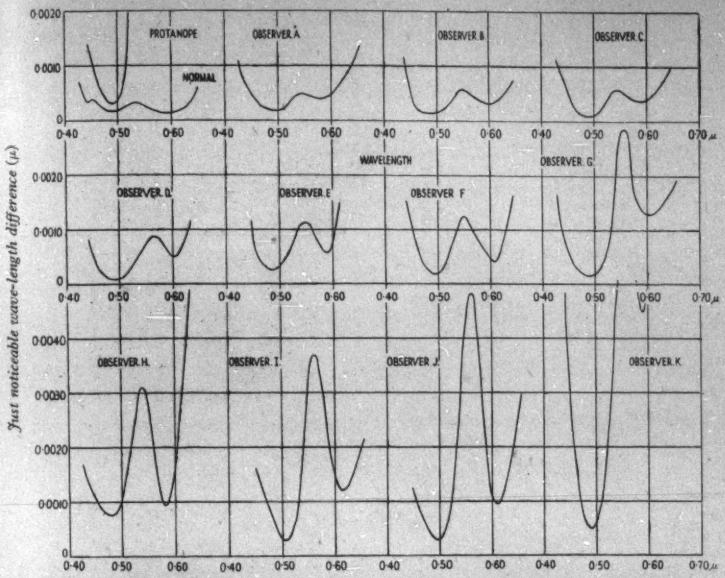


Figure 6. **Hue-discrimination curves for 11 protanomalous observers arranged in order of increasing magnitude of defect. The first diagram shows the corresponding curves for the normal trichromat and for the protanopic type of dichromat (McKeon and Wright, 1940).**

tendency is progressively poorer discrimination in the greenish-yellow. The luminosity curve of the deuteranomalous trichromats is similar to the deuteranopic curve.

Relation between dichromatic and trichromatic colours

The fact that dichromats can see all their colours as spectral colours has already been mentioned, and it is possible to show how the dichromat sees his colours in relation to the trichromat. In figure 8 is plotted the colour triangle for a normal observer, using red, green and blue primaries. The standard white is plotted in this diagram as a star, W. The dichromat sees white as a spectral colour, which is called the neutral point, and which is indicated in figures 2 and 5. For the average deuteranope this is at $\lambda = 0{\cdot}500\,\mu$. Therefore if the dichromat mixes a standard white and the colour at $\lambda = 0{\cdot}500\,\mu$, he will get a series of whites which will differ only in intensity. Therefore the straight

line AW, in figure 8, which may be continued to point C, represents a locus of dichromatic white points. The point C gives us the equation

$$0{\cdot}500\,\mu = W = 0{\cdot}68R + 0{\cdot}32B, \qquad \ldots\ldots(1)$$

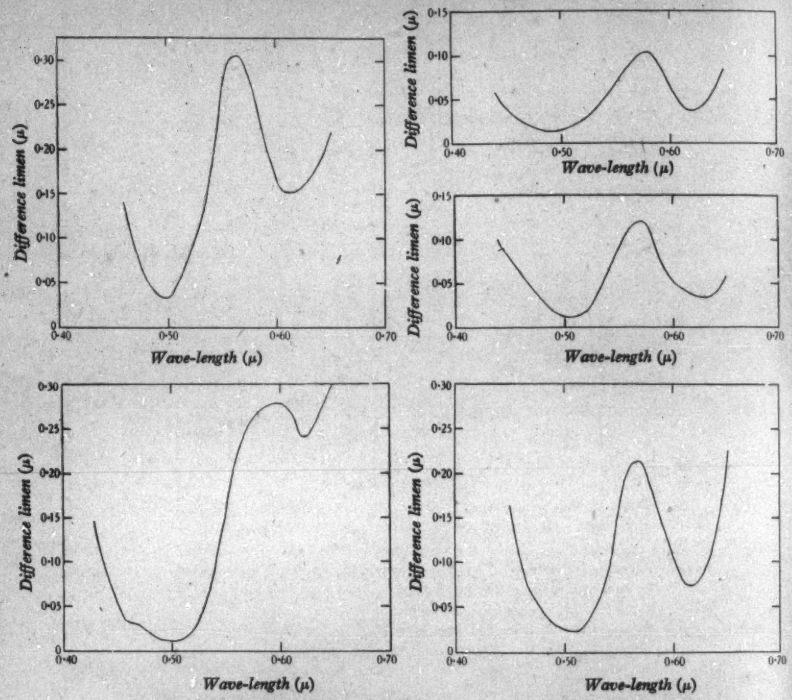


Figure 7. Hue-discrimination curves for 5 deuteranomalous observers (Nelson, 1938).

but from actual dichromatic matches we know that in terms of equal but arbitrary luminosity units

$$500\,\mu = W = 32{\cdot}26R + 1{\cdot}104B. \qquad \ldots\ldots(2)$$

Therefore, if we multiply the number of units of B in equation (2) by 13·75, we obtain the same relation between the R's and B's of equations (1) and (2), and thus establish a direct connection.

From the deuteranopic hue-discrimination curve we know that the discrimination at $\lambda = 0{\cdot}50\,\mu$ is $0{\cdot}002\,\mu$, and consequently, from $0{\cdot}498\,\mu$ to $0{\cdot}500\,\mu$

will represent one colour step. From the matching data we know how the deuteranope matches the colour at $\lambda = 0{\cdot}498\,\mu$ in equal, but arbitrary, luminosity units, and by multiplying the B value by 13·75 we know the value of this colour in terms of B and R on the green zero. This is point D. Starting from a convenient point, i.e. the white point, the boundaries of all the colour steps in the deuteranopic spectrum can be found, and the loci of these wave-lengths calculated and plotted in the colour triangle as shown in figure 8. The zones

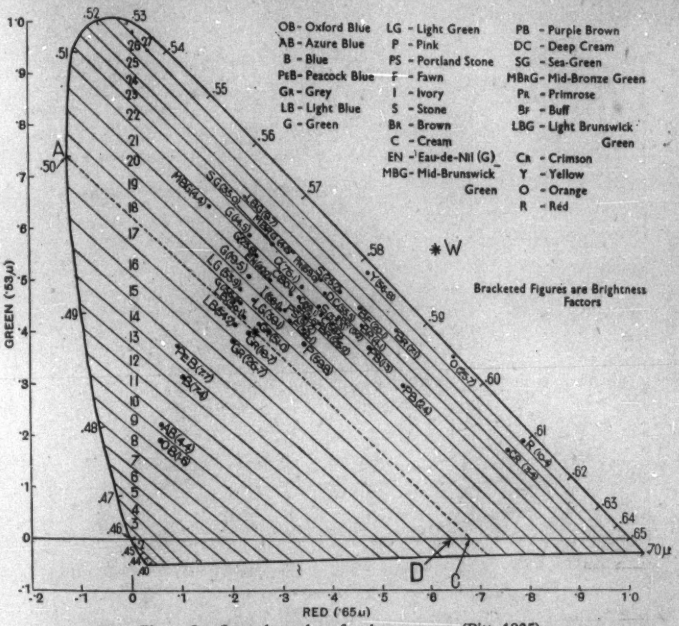


Figure 8. Iso-colour chart for deuteranopes (Pitt, 1935).

bounded by these lines contain, for the deuteranope, colours of the same hue. Figure 9 shows a similar diagram plotted for the protanope. Forty-three well-known paint colours, plotted in the diagram, indicate how the dichromats see these colours from the viewpoint of hue. It should be remembered, however, that the dichromat will only confuse colours in the same zone when the intensities, as seen by him, are of the same order.

Relation between normal and anomalous sensations

The colour triangle for the dichromat reduces to a straight line, and as it is known, from actual measurement, that the number of colours seen by the

anomalous trichromat is greater than the number seen by the dichromat, but less than the number seen by the normal trichromat, it is to be expected that the colour triangle, if plotted in exact relation to the normal colour triangle, should be smaller if the area is to represent the actual number of colours seen. If the colour triangles are plotted relative to the same physical stimuli it is obvious that these triangles will be of the same size, and it is therefore necessary to find some connection which will enable this triangle to be plotted in terms of the same subjective stimuli.

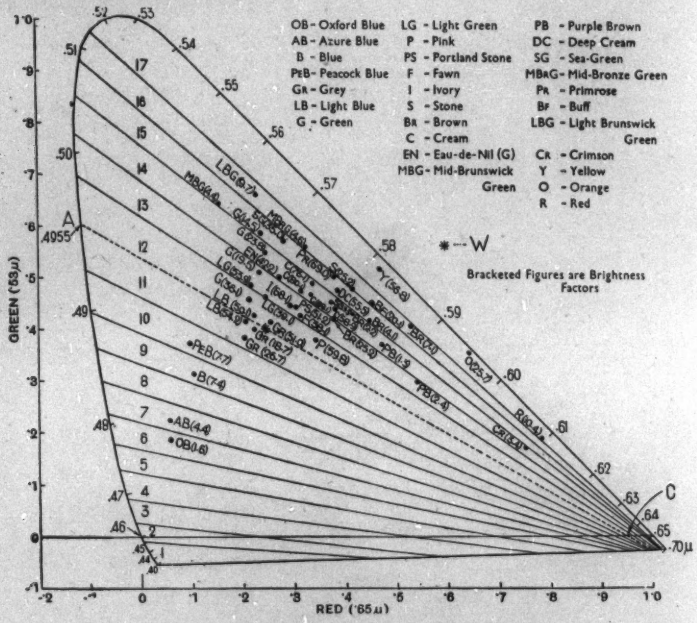


Figure 9. Iso-colour chart for protanopes (Pitt, 1935).

Nelson (1938) has discussed this possibility at length. He has assumed a previous suggestion put forward by Wright that the anomalous dichromat sees red and green as a desaturated orange and a yellow-green respectively. This is in agreement with hue-discrimination and saturation-discrimination data. From this it follows that the sensation of yellow is approximately the same for both the normal and anomalous trichromat. He further assumes, on the basis of McDougall's evolutionary theory, that the blue sensations are also the same in both cases. The points O and P, figure 10, represent blue and yellow

respectively, and are seen by both the anomalous and normal observers, subjectively as the same blue and yellow. If it is supposed that one hue-discrimination step produces the same change in colour both for the normal and for the anomalous trichromat it is justifiable to assume that the number of steps for the anomalous trichromat between yellow and red, at $0{\cdot}65\,\mu$, corresponds in colour sensation to a colour distant from the yellow by the same number of hue-discrimination steps as for the normal observer. In this way it is possible to find

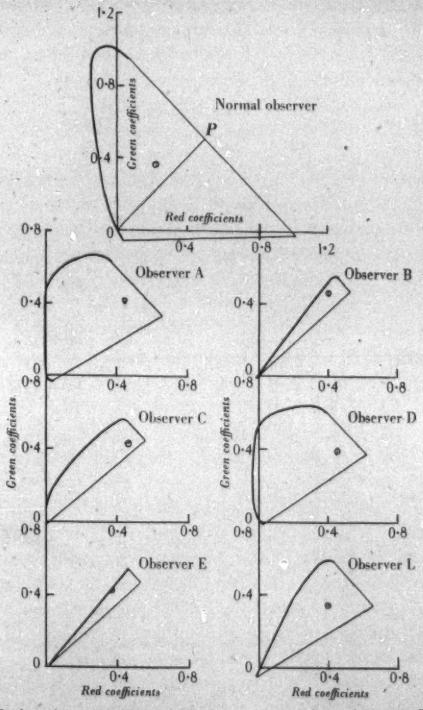


Figure 10. Reduced colour-triangles for 6 deuteranomalous observers (Nelson, 1938).

a point on the normal colour-locus which corresponds to the colour seen by the anomalous observer at $\lambda = 0{\cdot}65\,\mu$. In a similar but less justifiable way it is also possible to find another point on the normal colour-locus which corresponds to the colour seen by the anomalous observer at $\lambda = 0{\cdot}53\,\mu$.

In each case the number of colour steps between yellow ($\lambda = 0{\cdot}5825\,\mu$) and red ($\lambda = 0{\cdot}65\,\mu$) and yellow and green ($\lambda = 0{\cdot}53\,\mu$) were calculated from the individual

hue-discrimination curves and from the new coefficients for the red and green primaries, the coefficients of the blue primaries remaining constant; modified coefficients for the remaining wave-lengths were determined by ordinary transformation.

These modified coefficients are shown plotted in figure 10 and represent merely the colour triangle of the anomalous trichromat on a new set of primaries. Whether they represent, when viewed in relation with the colour triangle of the normal trichromat, a true subjective connection between the two systems is dependent on the validity of the assumptions made. Further evidence put forward by Nelson in some cases supports the results obtained, but in other cases disproves them.

The naming of colours

We do not know with any certainty what colours people with abnormal colour-vision see, but we do know that they see colours differently from the normal person. It is quite certain that the dichromat, with his limited range of colour and with the vast number of colour names at his disposal, makes use of different degrees of brightness to attempt to differentiate colours which to him

Table 1

Colour	Wave-length	Remarks by the dichromats				
		D_1	D_2	D_3	P_1	P_2
Red	0·6585–0·7800 μ	Yellow	Orange	Yellow-green	Dark colour —red ?	Orange
Orange	0·6000–0·6585	Yellow (warm)	Yellow	Orange-green	Yellow	Lemon
Golden-yellow	0·5828–0·6000	Yellow	Yellow	Orange-green	Yellow	Lemon
Yellow	0·5780–0·5828	Yellow	Yellow (orangy)	Orange-green	Yellow	Yellow
Greenish-yellow	0·5672–0·5780	Yellow	Yellow (orangy)	—	Yellow	Yellow
Green	0·5237–0·5672	Greenish-yellow	Yellow-white	Yellow	Yellow nearly gone, may be red ?	Light yellow to white
Blue-green	0·5020–0·5237	Whitish	Pink	Yellow	Yellow gone, nearly green	Blue
Blue	0·4306–0·5020	Blue	Mauve	Blue	Green	Royal blue
Violet	0·3900–0·4306	Blue	Blue	Blue	Dark colour	Violet

are the same hue. To a more limited degree the anomalous trichromat adopts the same procedure. Consequently the naming of colours under haphazard conditions can lead to erroneous conclusions. In an actual test a dichromat named various decreasing intensities of a green ($\lambda = 0{\cdot}53\,\mu$) as yellow, yellow,

reddish-yellow, red, red. The preceding table shows how various spectral colours were named by three deuteranopes and two protanopes.

Another interesting demonstration of the fact that colour-naming is a bad test was given by a protanope. He was shown a coloured plate in a book, containing three distinct colours, a red, green and yellow, and he was asked what colour the red was. His first answer was green, but within a second he had changed his mind and called it red. After his first answer he had allowed his eyes to wander on to the other colours, and although they look the same colour to him he could discriminate between them by means of their intensity difference. The brightest he called yellow, the next brightest green, and the faintest red. If it were possible for a dichromat (say a deuteranope) to name colours correctly, he should say, on being confronted with the spectral colour at $\lambda = 0{\cdot}51\,\mu$: "Among other colours it is a spectral green, a mid-Brunswick green, an eau-de-nil green, a cream, a pink, a brown, and a reddish purple" (see figure 7).

Although the naming of colours by persons with abnormal colour-vision is not so bad as would be expected from the hue-discrimination data, this is chiefly due to the remarkable faculty that they have of using the brightness property to estimate colour. Indeed, some anomalous trichromats spend a large part of their life without knowing their colour-vision is abnormal.

Qualitative tests for colour blindness

Various tests are used which have been designed to detect colour blindness. All these tests, with the exception of the Nagel Anomaloscope, present to the observer confusion colours chiefly in the form of reds and greens (see figures 8 and 9) of such intensity as to render these colours indistinguishable to the dichromat and barely distinguishable to the anomalous trichromat.

The Wool Test, originally designed by Holmgren, was improved upon by Abney. This improved apparatus has 6 test skeins and 30 confusion skeins, which are so mounted as to present only one test at a time. Used properly, this test usually detects the dichromat but seldom detects the anomalous trichromat.

The Stilling Charts consist of a number of irregular and vari-coloured dots printed in such a manner that certain numbers may be seen by the normal observer but not by the dichromats or extremely anomalous trichromats.

The *Ishihara Test*, which is really an improved form of Stilling Test, is designed so that coloured dots, forming numbers in the diagrams, are easily distinguishable by the normal observer, but cannot be distinguished or are misread by colour-blind persons. Some of the tests can be seen as numbers by the dichromats, but are not so apparent to the trichromats. Other cards include both the above tests. This test, which should be used under daylight illumination, is easy to use and quick in operation. It picks out dichromats with unfailing regularity, but despite claims to the contrary, does not always distinguish between protanopes and deuteranopes. It picks out the more pronounced of the anomalous trichromats.

In the *Board of Trade Lantern Test*, small coloured apertures are viewed at a distance of 20 feet and the examinee is required to name the colours. All the lights exhibited are of equal brightness to the normal eye, and as the luminosity curves of both the trichromats and deuteranopes are approximately equal, the latter are readily detected when greens and reds are present. Neutral filters, to alter brightness, are introduced when a suspected protanope or protanomalous observer is being tested. The test, which should be performed in a darkened room, approximates to field conditions and is used extensively for testing of personnel in the Royal and Merchant Navies, the railway companies, etc. A lantern, similar to the Board of Trade Lantern, has been described in detail by Martin (1939). This lantern, when correctly used, provides much information on the actual quantitative aspect of colour deficiency.

The *Nagel Anomaloscope* is an optical instrument which will detect and differentiate between protanopes and deuteranopes and between protanomalous and deuteranomalous trichromats. The observer sees a small field of view, divided into two parts, one half being composed of a variable mixture of red and green and the other half a yellow of variable intensity. Dichromats can match a full red or a full green with a yellow by adjusting the intensity of the yellow. The amount of red used will readily indicate whether the dichromat is a protanope or deuteranope (see luminosity curves, figure 2).

The anomalous trichromat will match the yellow with either a reddish yellow or a greenish yellow, according to whether he is protanomalous or deuteranomalous, and the actual proportions of red and green will be an indication of the type and amount of abnormality. While the anomaloscope may be regarded as one of the best tests for colour vision, it suffers from the disadvantage that a certain amount of skill is required on the part of the observers.

Colour in industry

From what has already been said, it is obvious that in any trade or profession where colour discrimination or measurement is a part of his duty, the colour-blind person may be a perfect example of a square peg in a round hole. A product may be ruined by the employment of such a person, and in certain cases, which will be indicated later, lives may be endangered. Statistical measurements

Table 2. Results of statistical survey (males)

	Nelson	Schmidt	Waaler	v. Planta
Total number	1338	6863	9049	2000
Protanopes (per cent)	1·27	1·09	0·88	1·60
Deuteranopes (per cent)	1·20	1·97	1·03	1·50
Protanomalous (per cent)	1·27	0·68	1·04	0·60
Deuteranomalous (per cent)	5·08	4·01	5·06	4·25
Total deficients (per cent)	8·82	7·75	8·01	7·95

show that about 8 per cent of the male population are in some way colour-deficient. This is shown in detail in table 2, due to Nelson (1938).

Less than one in ten of the colour-deficient are female.

A general indication of the types of difficulties against which the person with abnormal colour-vision is always striving are given below:—

A car driver, although not bothered by the ordinary coloured traffic lights, could not distinguish between the green and red lanthorns used for controlling traffic when road repairing is in progress (see figures 8 and 9). He always had to wait for some other car to come along so that he had someone to guide him through. The reason why he could distinguish the ordinary traffic lights was because he knew which position the " stop " and " go " lights appear, and also because the so-called green light is really a blue-green. He would be further helped if, instead of the three lamps being circular, they were, say, square, triangular and circular.

Another case concerns the unfortunate position of a physics graduate when he was taking his intermediate examination in chemistry. He was performing a titration experiment in which methyl orange was used as an indicator. To the trichromat, methyl orange is orange in alkaline solution and turns red when the solution is made just acid. To the dichromat it does not change colour in any way (see figures 8 and 9). I remember the vivid description given me by the person concerned of the relief he felt on passing this examination and finally severing his connection with methyl orange. Another example is more amusing than distressing. The person concerned stated that he could distinguish green cherries on a tree but he could not see red ones. He was a protanope, and the red cherries appeared almost black to him, and in the dark shadows of the leaves were almost invisible (see figure 3). The instance of a mourner appearing at a funeral wearing a red tie is well known. A copy of figure 3 would absolve him from all blame.

It is difficult to make a full list of all trades and professions and state specifically where colour blindness would prove detrimental. It is possible, however, to make reference to some of them, which may serve as an indication for all. Personnel for the Royal and Merchant Navies are subjected to very stringent tests. This is to enable those responsible to ensure that men who are colour-deficient are not placed in positions which would render them potential sources of danger. The look-out man must be able to differentiate navigation lights up to two miles in clear weather, or at safe distances in haze, rain, fog or snow. He must be able to distinguish between these lights solely by their colour quality. At sea, the dichromat as a look-out man would be dangerous, but it should be remembered that the anomalous trichromat, who can often differentiate between the colours when given time to deliberate, would cause delays which again might be dangerous. It is stated (Martin, 1933) that the anomalous trichromats present greater danger, due to the fact that they are more difficult to detect. For similar reasons, personnel for the railway companies are also

rigidly tested. Those who are colour-deficient are not employed on duties which necessitate the reading of coloured signals.

Soldiers employed in the Signal Corps and certain members of the R.A.F. are obvious cases where only normal trichromats should be used. The latter make use of coloured lights in one way or another, but it should also be realized that a pilot, when making a forced landing, has to make use of his ability to discriminate colours in order to ascertain that the particular spot on which he chooses to land is the safest one in the vicinity.

The four cases just described are ones where persons who cannot readily distinguish colour are a source of danger. In other industries or trades, persons may be required not only to distinguish colour but also to create and directly match them. These persons may not endanger life, but they would certainly prove to be a liability. The paint and varnish industry is a good example of such a trade. Mid-Brunswick green and some types of brown paint are of almost equal brightness to the dichromat. He sees them as the same colour (figures 8 and 9) and would match them without hesitation. It is not beyond the bounds of possibility for him to match an actual crimson paint with a light Brunswick green. Such glaring examples are merely instances of the capabilities of the dichromat and extreme anomalous dichromats, and would probably be readily detected. It is when paints of smaller differences are matched (and here the majority of the anomalous trichromats enter the field) that the colour deficient prove themselves a real nuisance. It goes without saying that such people employed as colorimetrists are worse than useless. From the colour viewpoint, that which applies to the paint industry similarly applies to other such industries as woollens, textiles, dyes and inks, and also to persons retailing these products.

A special mention should be made of the paper and other industries where the measurement of whites are of utmost importance. It has been stated that measurements of the colour coefficients of whites should be made to an accuracy of at least one in the third decimal place, to satisfy the requirements of the users. The dichromat would distinguish himself by getting the first decimal place wrong and even then not being consistent, while the anomalous dichromat would probably get the second decimal place wrong. Even persons with a slight colour deficiency would be incapable of making a correct normal reading. It has been generally realized by those directing industries where extreme accuracy of colour matching or measurement is required, that only observers possessing closely normal colour-vision should be employed in certain positions. Excessive variation of the macular pigment from the normal would also introduce errors too large to be acceptable, and tests which measure the Y.B. ratio have been devised with the view to detecting persons who, although possessing normal trichromat vision, suffer from this defect. This test, suggested by Ives and Kingsbury (1915), consists in comparing the amount of light transmitted by two filters, one yellow and the other blue. The ratio of the brightnesses for the

normal eye can be calculated from a knowledge of the colour temperature of the source and the spectro-photometric curves of the filters. The extent of the departure from this ratio indicates the degree of abnormality of the observer. In fairness to all these people, however, it should be placed on record that spectrophotometric readings, being purely physical measurements, could be correctly made by all classes of colour-blind observers. Even so, it should be noted that, due to the low luminosity in the red, the protanopes and protanomalous trichromats would not be able to measure the intensity of the longer wave-lengths with any degree of accuracy.

The manufacture of cosmetics has recently developed into something approaching a major industry. Women, with their very keen sense of discriminating even the most subtle difference in colour, would be very quick to notice a coloured preparation created by a colour-deficient person.

The ability to discriminate small differences in colour is often of great value to the practical physicist. He would find it quite impossible to determine the colour temperature of a lamp visually with the required degree of accuracy if his colour vision was in any way abnormal, because his discrimination of colours would not be good enough. Manufacturers of colour-temperature meters, the use of which depends on the blending of a green and yellow to match the colour of the subject, state specifically that the apparatus should not be used by colour-blind observers. As users of a flicker photometer, for the measurement of heterochromatic colours, their accuracy as compared with the standard observer would depend upon whether they were protanopic or deuteranopic. If of the first class, whether a dichromat or anomalous dichromat, an observer would give erroneous readings, the magnitude error of which would depend on the nature of colours being matched. This error would, in the case of the protanomalous worker, also depend on the extent of his colour deficiency. The error may be calculated from the luminosity curve of the actual observer and the spectrophotometric curves of the colours to be matched (if known). Due to the luminosity curves of the deuteranopic and deuteranomalous observers being similar to the standard curve, the differences in these cases are not so marked. The extent to which colour is important to the spectroscopist is obvious. The method of measurement adopted and the great use made of photography protect him to a large extent if he is colour-deficient. Colour-staining of microscopic sections can be misleading to the colour-deficient microscopist if the colours produced are red, yellow, or green.

The use of methyl orange as an indicator has already been mentioned. The chemist also makes use of flame tests which indicate to him the nature of the substance under test, and many chemical reactions are characterized by a change in colour, often a small one. While it would be foolish to assume that a colour-blind observer could not follow the profession of chemistry or physics, it is certain that his usefulness is restricted in certain branches of these subjects, and awareness of the nature and extent of the defect would be of use to him.

What has been said about the chemist also applies, although to a lesser extent, to the pharmacist. Members of the medical profession also rely to a degree, which is probably greater than they realize, on colour as an aid to diagnosis.

The photographic industry provides another outlook on the subject. Many of the sensitized goods have to be processed in red light. Protanopes and the more extreme cases of protanomalous trichromats would find it almost impossible to work under such conditions. It should be realized, however, that colour-blind persons would have no difficulty in working under panchromatic (green) safe-lighting conditions, as none of these persons, not even the so-called green-blind, have luminosity deficiency in the green. The obvious conclusion from the above remarks is that only protanopes and protanomalous trichromats need be debarred from working in dimly illuminated darkrooms, and only then when red lighting is employed. The preparation of colour films is another concern of this industry. As it is impossible, for reasons too involved to be discussed in this paper, to reproduce actual colours accurately, it is necessary to reproduce them in such a way that the complete picture, as a whole, looks correct. Although the individual colours may depart markedly from the original colours, and the whole still be accepted as a good reproduction, what may be termed the balance of colour must be fairly exact to obtain the effect. Many tests and much checking against standards must be effected by competent personnel before high-class colour pictures are obtained, and it is evident that the colour-vision of this personnel must be normal in every respect.

What has already been said about colour-films also applies to the work of the interior decoration artist. This artist must be able to present a good "colour scheme", and while it is not everyone who could follow such an occupation with success, it is obvious that the colour-deficient person would never be able to perform such duties successfully except for persons of the same degree of colour deficiency as himself. It is here that colour is really treated subjectively, being regarded in contrast with other colours, and many of the industries described must employ persons capable of treating colour in this manner. For instance, the person creating a primrose-yellow paint must produce a paint the colour of which resembles the colour of the primrose when seen in contrast with its dark green leaves. In fact he produces, objectively, a more saturated yellow which, when seen on a wall and without the contrasting green, looks like the yellow of the primrose.

It is not only in the trades directly concerned in the manufacture of coloured goods that persons with normal colour-vision should be employed (or rather, from which persons with abnormal vision should be excluded). Most people have passed pleasant times watching workmen sitting in a small hole in the pavement quickly and nonchalantly joining thousands of telephone wires together, and have marvelled how they managed it. The telephone cables, sometimes containing as many as two thousand pairs of cables, are colour-coded. The pilot wire, from which the start is made, is usually coloured orange, and spiralling

from the pilot are the other wires, which are coloured somewhat differently in the different-sized cables. In one of the smaller types of cable, the colours are white, black, blue, green, red, red and black, blue and black, green and black. It is known that colour-deficient persons have difficulty in joining these cables, but providing they keep the order of wires correctly they can accomplish the jointing. With the larger cables, which contain a more complicated colour-system, they find it correspondingly more difficult, and often, when trying to locate a single pair of wires on which a fault has occurred, they find the task beyond their capabilities and are forced to seek advice from the mate who is invariably helping in the operation. The obvious moral is that if the jointer is colour blind, his mate must have normal colour-vision.

Many factories have colour-coded supply systems. Thus A.C. electricity, D.C. electricity, hot and cold water, gas, steam, brine and air are housed in pipes marked in some way with a distinguishing colour. Many of these pipes are similar in size, and while it is not suggested that the colour-deficient person would unwittingly connect, say, the gas and water together, it might occur that, after cutting off the steam system in order to repair it, he would then proceed to take the water pipes apart to the disadvantage of all concerned. What is probably more important is that other minor troubles, not always brought to the notice of the authorities, are almost certain to occur which cause delay of some description.

It is hoped that, from what has been said on the subject, the need for testing the colour vision of all persons connected with certain trades or professions will be realized. It is, however, obviously to be desired that the tests should not be made after the person has commenced employment, as his subsequent removal might lead to considerable economic hardship. The tests might be made before entering a given occupation or, better still, before the future employees have left school. This would then allow time for a more appropriate choice of employment to be made; it might prevent training being given that subsequently turned out to be useless; it might save the child a certain amount of ridicule, embarrassment and difficulty while still at school. The tests would, presumably, have to be made by the medical officer, but it should not be overlooked that the science master or the art master would almost certainly be interested in the test, and his co-operation might help considerably to reduce the labour. In general, the tests of the Ishihara type are best where mass testing is required. Where the defect is likely to cause danger, a further test must be adopted to ensure that all the colour-deficient persons are weeded out; for this, the Board of Trade lantern test is usually used. Personssuspected of being colour deficient should, if possible, be further tested to ascertain whether they are dichromats or anomalous trichromats, and whether they are of the deuteranopic or protanopic types, for often (as in the case of deuteranopes and deuteranomalous trichromats in red-illuminated darkrooms) some classes of colour-deficient persons can work with the same efficiency as normal trichromats even when colour plays an

important part in the scheme of things. This dividing of colour-deficient persons into their various groups can only be performed with any degree of certainty by using an apparatus of the Nagel-anomaloscope type. As already stated, this apparatus in its present form is difficult for an unskilled observer to use, and its re-design to obviate this rather formidable difficulty should be one of the aims of those responsible for the testing of colour blindness. It is further hoped that one day the various responsible authorities will cease to colour-code in the so-called confusion colours, as in the green and red lanthorns of the road repairer, and will do more to mitigate the lot of the colour deficient, in the same way as a golfer acquaintance, a protanope, who realized that, because of his low luminosity in the red, he could not see red golf-pegs, and consequently lost them, solved his difficulty by using yellow ones.

REFERENCES

Abney, W. de W., 1913. *Researches in Colour Vision and Trichromatic Theory* (London).
Gibson, K. S. and Tyndall, E. P. T., 1923. *Sci. Pap. Bur. Stand., Wash.*, no. 475.
Guild, J., 1924–5. *Trans. Opt. Soc.* **33**, 73.
Guild, J., 1931. *Philos. Trans.* A, **230**, 49.
Hecht, S. and Schlaer, S., 1936. *Amer. J. Physiol.* **20**, 57.
Ives, H. E. and Kingsbury, E. F., 1915. *Trans. Illum. Engng Soc., N.Y.*, **10**, 203.
Jones, L. A., 1917. *J. Opt. Soc. Amer.* **1**, 63.
Konig, A., 1903. *Gesammelte Abhandlungen zur physiologischen Optik* (Leipzig).
Konig, A. and Dieterici, C., 1892. *Z. Psychol. Physiol. Sinnesorg.* **4**, 241.
Laurens, H. and Hamilton, W. F., 1923. *Amer. J. Physiol.* **65**, 547.
Martin, L. C., 1933. *Spec. Rep. Ser.* (Med. Res. Counc., Lond.), no. 185.
Martin, L. C., 1939. *Brit. J. Ophthal.*
McKeon, W. M. and Wright, W. D., 1940. *Proc. Phys. Soc.* **52**, 464.
Nelson, J. H., 1938. *Proc. Phys. Soc.* **50**, 661.
Pitt, F. H. G., 1935. *Spec. Rep. Ser.* (Med. Res. Coun., Lond.), no. 200.
Steinder, O., 1936. *Sitzungsber. Akad. Wiss. Wien*, **115**, 39.
Wright, W. D., 1927-8. *Trans. Opt. Soc.* **29**, 225.
Wright, W. D., 1929. *Spec. Rep. Ser.* (Med. Res. Coun., Lond.), no. 139.
Wright, W. D. and Pitt, F. H. G., 1934. *Proc. Phys. Soc.* **46**, 459.
Wright, W. D. and Pitt, F. H. G., 1935. *Proc. Phys. Soc.* **47**, 205.

DISCUSSION

Dr. Shaxby. I am glad that Dr. Pitt treated the subject from what he called the physical standpoint, i.e. of the experimental investigation of the facts of colour vision and its abnormalities; theories of colour vision are not yet such as to help us in the practical questions of the disabilities for various professions that we are considering to-day. We still require more statistics, and particularly a good quick method of assessing trichromatic anomalies. Colonel Ishihara's test still seems the most satisfactory preliminary one for general use.

I think the Colour Group can serve a very useful purpose in industry: we know all too little of the occupations in which colour deficiencies are a bar, or at least a handicap, by slowing up the work.

To the professions Dr. Pitt listed we may add that of the geologist, with his special maps, coloured to indicate the various strata.

I consider that it is very desirable that tests of colour vision should be made in schools, early enough to prevent boys preparing for occupations for which colour blindness will unfit them, as soon, in fact, as they are of age to make the tests satisfactorily. In secondary schools I think the test should be made at the time of the initial medical examination; in primary schools not later than at the last medical examination. In one school alone, in which I am interested, in the last few terms two boys have been rejected for the R.A.F. and one for the Merchant Service; not one of them had any suspicion that his colour vision was not normal.

I hope that a special sub-committee of the Colour Group may be formed, whose duties shall include, among other things, the obtaining (from the various industries interested in questions of colour) of information, as complete as possible, of the techniques and processes in which colour deficiency is a bar, and particularly those in which it causes slowness or uncertainty.

Dr. C. Maunder Foster. The accuracy of colour vision is to some extent a function of training, and the degree of colour blindness might be more accurately measured than at present; standard limits might be established between which colour perception might be considered adequate for various purposes. The marked discrepancy in relative incidence of colour blindness between male and female may indicate some connection between the gonadotropic hormones and the mechanism of colour vision.

The accurate use of indicators in analytical work is an instance of an important case where slight errors in colour vision gave markedly different results. I should like to ask whether slight errors in colour vision could in some measure be compensated by altering either the colour temperature of the illumination or by filters.

Dr. Strange. I wish to support the suggestion that there should be a regular colour-vision test of all children. The need for this test is usually underestimated because of our feeling that any anomalies in vision can hardly pass unnoticed, at least among scientific workers. This view ignores how essentially normal even a dichromat's colour perception is to himself, and it is only unusual circumstances, such as green berries on a holly tree, or special tests, such as indicator end-points, that disclose any abnormality. This may be illustrated by my experience among one group of about 50 scientific workers, where three hitherto unestablished, if not unsuspected, colour-blind persons were discovered. The first two were anomalous trichromats, one a physicist of over 30 and the other a laboratory assistant; the third was almost a dichromat, and had been given the job of testing cathode-ray tubes, including the colour of the screens!

These examples indicate another essential feature of the proposed tests,

It is not enough to sort children out into those having standard and anomalous vision; subsidiary tests must be established to define the degree of anomaly and limits set up to indicate what degrees are likely to be a serious handicap in different occupations. In two of the cases mentioned, the anomalous vision caused only slight inconvenience, and was in no sense a bar to success in the chosen occupations.

Dr. WALSH. Is the Ishihara test sufficient as an indication of suitability or unsuitability for such occupations as the Navy or the railway service, or would it not be necessary to use a lantern test as well?

Mr. PRESTON. I would like first to ask a question, to make certain that I have properly understood part of Dr. Pitt's very interesting discourse. Is it correct that, among defective observers, the poorer are the powers of hue discrimination, the farther the luminosity curve is displaced from the normal or average? In other words, is it possible for some observers to have a luminosity curve approximating to the average or "standard" eye and yet be deficient as regards recognition of colours?

I suppose it is difficult or impossible at the present stage to indicate what are the absolute, as distinct from relative, sensations produced by light of different wave-lengths for defective eyes. The curves of luminosity function on a *relative* basis show the differences in sensitivity to radiation of different wave-lengths for the same eye, but do not necessarily provide data for comparing the brightness sensations produced in two *different* eyes by the *same* monochromatic source.

Dr. Pitt has hinted that the low values of the relative luminosity function for the protanope in the red indicate a low *absolute* sensitivity, since the protanope is unable to see at all by the light of a red photographic dark-room lamp. This suggests that curves of threshold values of illumination, or, rather, radiation, throughout the spectrum would be interesting; but perhaps I am anticipating too much in this remark.

Dr. W. D. WRIGHT. I should like to support Dr. Pitt in his belief that in many industries colour blindness must be a handicap both to the individual and to the firm concerned, and I think there can be no doubt that testing for colour blindness in such industries is very much to be desired. We have to remember, however, that those people who were found to have defective colour-vision might be exposed to considerable economic hardship if they were shown to be unfitted for their particular job; and any campaign to encourage testing in industry would have to be accompanied by assurances that there are a great many activities where colour blindness is no bar to efficiency and to which defective observers could be transferred. From this point of view, the suggestion that tests should be made in schools is excellent, since it would sort out the colour-defective cases before they entered industry and would avoid any suggestion of subsequent hardship.

There are a number of practical questions that arise in connection with testing at schools, and I have put certain of these to Dr. Gale, of the Board of Education, as he was unable to be present when Dr. Pitt read his paper. They are dealt with in the following contribution, but Dr. Gale has asked me to say that any opinions expressed are his own and do not necessarily represent those of his Department.

On one point, Dr. Pitt's paper is slightly misleading because the information is incomplete. Physicists generally simplify colour problems by reducing them to two variables instead of three, by regarding brightness differences as being distinct from colour differences. Thus, figures 8, 9 and 10 are really plane sections through the colour solid, and when Dr. Pitt says that a trichromat can distinguish about 5000 colours with a 2° field, this refers only to the plane diagram and not to the colour solid. From the point of view of the dyer, the painter and the man in the street, a brown is a different colour from a yellow or orange, although they may be represented by the same point in the colour triangle, and when brightness differences are included, the number of distinct colours runs into millions. To get a complete picture of the colour perception of the colour-blind observer, we ought, therefore, to consider what happens to the trichromatic colour solid (as described, for instance, by Newhall (1939)*) in the case of the anomalous trichromat and dichromat. Presumably, in the case of the anomalous observer, the solid is reduced in volume in various directions according to the type of defect, while the dichromatic solid becomes a plane. Has Dr. Pitt any information on this question?

Dr. GALE. I should like to submit the following notes in reply to questions which Dr. Wright has sent to me:—

1. In 1938, 1518 doctors were employed on school medical work. The majority of these did other work, usually in some other branch of the Public Health Service, as well, and the total amount of time which they gave to school medical work was equivalent to that of about 728 whole-time doctors. This was about 1 to 6220 children in average attendance.

2. Medical inspections are divided into (*a*) routine, (*b*) special—when the child is brought forward for examination by parent, teacher or nurse, for some special reason, (*c*) re-examination of children found at a previous inspection to need special care. The total numbers of inspections in 1938 were:—

Routines	1,677,008 (37 % of total number in average attendance)
Specials	1,563,917
Re-examinations . . .	2,182,157

* S. M. Newhall, 1939. *Amer. J. Psychol.* **52**, 383.

3. In elementary schools, routine inspections are usually carried out on entrance (about age 5), at about age 8, and at about age 12 (some areas have four inspections). In secondary schools, procedure varies considerably, but generally four full routine inspections are done during a child's school career. The conditions of working vary from a screened-off corner in a classroom of a village school (with lamps as the only artificial light) to a well-appointed inspection room in a modern school. Usually about 5 to 8 minutes are allotted to each child. This time varies very much from child to child, but if more detailed examination is necessary the child is usually referred to a clinic. Vision is often tested beforehand by the school nurse, using Snellen's types, and she mentions to the doctor any child who appears to have a defect. Vision is usually tested only at the two later inspections because of the difficulty and unreliability of testing young children who do not know their letters.

Now for the answers to particular questions.

(1) *Would the introduction of a colour-vision test be feasible?* I think the answer to this depends very largely on what the test is. It would have to be quick, not too complicated, and capable of being carried out in an ordinary room, possibly without electric light. This is where the Colour Group could be most helpful, i.e. in the choice of possible tests.

(2) *Would it be better to test elementary or secondary school children, or both?* If a test is thought desirable, I think it should be carried out for (*a*) secondary-school children at their last examination, and (*b*) elementary-school children, also at their last examination. The answer to this question does, however, depend on a good many things, and I should not like you to regard it as anything but tentative.

(3) *Would it be the responsibility of the medical officer or the teacher to advise on suitable employment?* I do not think that the medical officer could know enough about industry to decide how far a defect would interfere. It is part of the medical officer's duty to give general advice on employment at the time of the last medical inspection. His advice is usually with reference to things like unfitness for heavy physical work or unfitness for work in a dusty atmosphere, etc. Such advice is, of course, given in consultation with the head teacher.

(4) *Would the Board approve of testing by the science master?* I do not think I can answer this, as it raises some rather thorny questions. From the purely medical point of view, I see no objection to the test being done by anybody who is competent. It must be remembered that there is no special science master in very many elementary schools.

(5) *Vocational guidance.* Neither the school medical officer nor the teacher could be expected to know the exact requirements of a given industry. I should have thought that the certifying factory surgeon was the man to deal with

this aspect. Another question which arises is—in how many jobs is a defect of colour-vision important, i.e., would it be justifiable to test the colour-vision of something over half a million children a year if only, say, 1 % or 10 %, or whatever the percentage is, were going to jobs where colour-vision mattered? Some estimate of the percentage of children entering such industries would obviously be important, and I confess I have not the remotest idea of what it really is. Here, again, I am sure members of the Colour Group would have useful information.

(6) *A.T.C. admission.* This is really a matter for the Air Ministry, and I am speaking rather without the book, but I gather that a defect of colour-vision is no bar to admission, because there are many jobs in the Royal Air Force where the defect does not matter. When candidates are transferred from the A.T.C. to the R.A.F. they go through the full R.A.F. medical examination, which, I believe, includes tests of colour-vision.

(7) If representations were made to the Board of Education I am sure they would be carefully considered, but I think an overwhelmingly strong case would have to be made out before anything new could be added to the work of school doctors, or teachers, in present circumstances.

Prof. L. C. Martin. I approve the suggestion that tests of colour vision in schools should be instituted when opportunity offers, but it will not be a simple matter to arrange the scheme. For example, it will not be sufficient to tell a lad: "You are an anomalous trichromat, and therefore cannot hope to undertake watch-keeping duties in the Navy." There must be an opportunity for a supplementary examination if required—and this supplementary examination would possibly have to be of a more or less official character.

But even under war conditions, some colour-vision testing might well be undertaken in secondary schools as a voluntary measure, and would tend to prevent disappointment while helping to make sure that effort will not be misdirected in preparation for forms of national service for which an individual is not suitable.

Author's reply. There is little to be said in reply to Dr. Shaxby beyond supporting him in his plea for the more intensive testing of school children and to say that the sub-committee to which he refers is already taking shape.

In reply to Dr. C. A. Maunder Foster, slight errors in colour vision can be compensated by the use of suitable filters. This is described in detail in the paper by Ives and Kingsbury (1915), to which I have already referred.

Dr. Strange's discussion emphasizes the need, which in my opinion is real, for adequate as opposed to superficial colour-testing.

In reply to Dr. Walsh, the Ishihara test is not considered sufficient as an indication of the suitability for such occupations as the Navy or the railway service. A lantern test is always used in addition to the Ishihara test.

In reply to Mr. Preston, there is no evidence whatever for assuming that poorer powers of hue discrimination are in any way connected with the displacement of the luminosity-curve. The deuteranopic luminosity-curve is very similar to the normal curve, and yet the discriminatory power of the deuteranope is much lower than that of the normal observer. Figure 6 shows an increasing magnitude of defect in the discriminatory powers of 11 protanomalous observers, but these observers have almost identical luminosity-characteristics. The same phenomenon is also shown by deuteranomalous observers. It is quite impossible to measure the absolute sensation curves of the eye, but by making various assumptions it is possible to indicate the possible shape of such curves. There is no doubt that the protanope has a low absolute sensitivity in the red, and that threshold measurements would supply interesting and useful data. To what extent these data would be used, in comparison with similar data completed for normal observers, is difficult to foresee, as once again " various assumptions " would have to be made.

I find Prof. Martin's query, as to what is to be done for a young person after he is told that he is colour deficient, difficult to answer. All I can suggest is that he should be told by a competent person what particular jobs should be avoided. As pointed out, it does not always follow that he should avoid occupations where use is made of colour.

The first part of Dr. Wright's discussion may be regarded as a partial answer to Prof. Martin's query. With reference to the colour solids, for a proper understanding of what I consider to be a difficult aspect of the subject, the paper, which took longer than the time limit set to deliver, would have tended to become interminable. I agree, however, that only the colour solid presents the true picture, but beyond agreeing with the assumptions which Dr. Wright makes, I have nothing to add beyond expressing the hope that one day somebody will perform the useful but arduous task of expressing the characteristics of the anomalous observers in this form.

Dr. Gale's reply to the questionnaire put to him by Dr. Wright forms a very useful basis on which to start on the problem of testing school children. Dealing with a query raised by Dr. Gale, there is no doubt that the Ishihara test fulfils his requirement of a speedy test capable of being used without complicated accessory equipment. A person considered competent to teach should be able to supervise the use of these tests. The question of vocational guidance is again raised, and it would seem that the solution of this problem will be an important task for the aforementioned sub-committee.

CANAL-RAY POSITIVE-ION SOURCES

By J. D. CRAGGS, M.Sc., A.Inst.P.,
Metropolitan-Vickers Electrical Co., Ltd.

Received 8 *January* 1942

ABSTRACT. The paper deals with the positive-ion output available from cold-cathode glow discharge tubes with a canal drilled through the cathode.

The importance of obtaining large canal-ray currents for high-voltage positive-ion bombardment work is stressed, and a brief preliminary discussion of previous work given. The reasons for adopting canal-ray sources are then discussed. A study of the discharge mechanism of such sources reveals several interesting features.

In the measurement of intense positive-ion currents, precautions were taken to avoid errors due to secondary emission of electrons from the collector, and, a satisfactory technique having been developed, exhaustive measurements were taken on canal-ray sources of two types differing in detail from each other.

The necessity for using a low gas output is stressed, and the effect on yield of varying the canal-flow impedance is studied. Some interesting results were obtained by the use of a focussing diaphragm in the source, e.g. yields were considerably increased, but with a higher tube pressure. Total ion currents up to 2 ma. in strength were obtained.

The presence in the canal-ray beam of large numbers of neutral particles was shown by an experiment on electrostatic deflection of the rays after leaving the canal.

The final part of the paper deals with magnetic deflection of the beam and its analysis into different groups of ions. This work was done with a high-voltage cascade acceleration tube, and qualitative data on collision processes in molecular hydrogen are given.

§ 1. INTRODUCTION

The production of focussed positive-ion beams of high current-density ($\sim$100 microamps/cm^2) is now a problem of extreme importance in the technique of artificial nuclear disintegrations. The inefficiency of such reactions must be offset as largely as possible by the use of large numbers of bombarding particles. For instance, 50 μa. of protons at 550 kv. cause an emission, from lithium targets, of gamma rays equivalent approximately to 0·5 mg. Ra (a yield of the order $7/10^{10}$ photons/proton). If such gamma rays are now used to cause neutron emission from a secondary target, there is a further large diminution of effective output, and the results are positive in these conditions for only a few substances (Bothe and Gentner, 1939; Chang, Goldhaber and Sagane, 1937). It is therefore desirable to produce focussed ion beams of 0·1 to 1 ma. at medium voltages ($\sim$500 kv.). As nearly all ion sources produce strongly divergent beams, the focussing problem is largely an independent one, and will be considered separately.

It is not difficult to produce a source yielding 1 to 100 μa. of positive ions, and almost any gaseous discharge device with a perforated cathode will give emergent currents of this order. The problem of increasing the output is not entirely one of increasing the power input to the source. Some of the difficulties are discussed below, with reference particularly to the type described by Oliphant and Rutherford (1933), i.e. the low-pressure, high-voltage, glow discharge-tube with a cathodic canal.

§ 2. TYPES OF ION SOURCES

With the exception of a few special devices, such as Franzini's source (1938), in which a palladium tube with an internal pressure of 1 atmosphere of hydrogen surrounded by a negatively charged electrode gave 100 to 300 μa. of positive ions (20 kv.) at about 10^{-5} mm. Hg pressure, positive-ion sources fall into two main classes: (*a*) low-voltage arcs, and (*b*) glow discharges. Before proceeding to the description of work carried out on sources of type (*b*) it is useful to draw comparisons.

Type (*a*), which is widely used in America, consists usually of an arc passing between a hot cathode (through a constriction) and a massive anode. The ions are drawn out through a hole leading radially into the constriction by means of a probe to which a relatively high potential ($\sim$5–10 kv.) is applied. In some tubes no probe is used, and ions and electrons diffuse out of a radial hole. The arc voltage may be 100 to 200 at a current of 1 to 2 amp., and a gas pressure (hydrogen or deuterium) of 0·01 mm. Hg. Some excellent results have been obtained with sources of this kind. Lamar, Buechner and Compton (1937), Timoshenko (1938) and Zinn (1937) obtained outputs of several hundred μa. at arc currents of 1 to 4 amp. ($\sim$100 v.). The capillary, through which the arc passed, was 3·5 mm. diameter.

In the case of the cold-cathode glow discharge sources, the experiments of Oliphant and Rutherford (1933) have formed the basis of all subsequent work. (See, for example, the experiments of Bouwers, Heyn and Kuntke, 1937.) Their source consisted essentially of two heavy concentric steel cylinders for the anode (inner) and cathode (outer), the latter being closed at one end with the cathode proper, which had a central hole bored in it. The cylinders, of length about 1 metre and diameter 12 cm., had a small sideways clearance (few mm.) to prevent stray discharges, the main discharge passing from the region inside the anode cylinder to the cathode block, from the hole in which the canal rays emerged. For 20 ma. discharge current at about 20 kv., a beam strength up to 1 ma. was obtained. It was stated that, at first, the beam consisted mainly of molecular ions but, after running, these broke up and the beam became composed principally of protons (hydrogen gas was used) and neutral particles. Sources of this type have become general in European practice. Bothe and Gentner (1937) and Amaldi *et al.* (1939) have used them, and obtained target currents of about 50 μa. in high-voltage acceleration tubes, i.e. their source output currents were probably about 100 μa.

Roberts (1937) has published a direct comparison between arc and canal-ray sources, also using a complete disintegration set. The canal-ray source gave 100 to 200 μa. output, but only 2 to 3 μa. of ions of mass one were focussed on the target. A high proportion, about 50%, of atomic ions was obtained, but it was considered that the power and voltage required were disadvantages. Oliphant (1935), comparing the two types of source, considered properly designed canal-ray tubes to be most suitable, and stated that arcs require, in general, hot filaments of short life, and pure gases. It also seems clear that arc sources are more complicated to build and operate than the glow-discharge tubes.

Finally, the comprehensive work of Hailer (1938) on canal-ray sources should be mentioned (see below for further references). The curves (loc. cit.) indicate an output current of about 1·5 ma. at 30 kv., 25 ma. discharge and a pressure of about 0·18 mm. Hg of hydrogen. The proton content was about 30%.

The above discussion ignores the problem of focussing the beam after it has left the canal. Both types of source give divergent beams, but as the glow-discharge sources in general run at a much higher voltage (20 to 50 kv.), correspondingly higher focussing voltages are required. This is not usually a serious defect, as such sources are almost invariably used with high-voltage (100 to 1000 kv.) apparatus, and potential division, either by corona points or resistances, is usually feasible.

§ 3. MECHANISM OF GLOW DISCHARGES AT LOW PRESSURES

In order to obtain as high a yield as possible from canal-ray sources it is necessary to consider the physical properties of the discharge. For the highest efficiency, the latter should be run near the low-pressure limit of such discharges, probably because the radial components of the high field-strength thereby encountered are effective in partially concentrating the positive-ion current before the cathode is reached. A larger proportion of the current is thus available for the formation of canal rays. It is found that the appearance of the discharge in these conditions is different from the normal régime, in that the various layers (cathode glow, cathode dark space, negative glow, negative dark space and positive column) largely disappear. Most of the tube becomes filled with the cathode dark space (Hailer, 1938), and the cathode glow extends outwards from the cathode and recedes from the tube walls, thus forming a thin pencil. Experiments, including some by the author, have clearly shown the continuous nature of the change in normal cylindrical tubes with flat electrodes. In general the sharpness of the pencil seems to be independent of tube geometry.

The luminosity of the cathode glow, and thus of the pencil ray, is due to ionization processes initiated by positive ions or neutral particles (collisions of the second kind). This is shown partly by the fact that the pencil passes through the canal without appreciable alteration in structure. The cathode glow is a region of positive space-charge. Laporte (1939) stated that its luminosity is due to excitation or ionization by electrons (collisions of the first kind), but was

probably confusing the cathode glow with the so-called first cathode layer which, at least in neon in certain conditions, is a separate phenomenon (Holst and Osterhuis, 1923; Seeman and Orban, 1935; Druyvesteyn and Penning, 1937).

Güntherschulze and Keller (1931) have clearly described the relation between the first cathode layer and cathode glow. The former is found (Ar, He, Kr, N_2 and O_2) for low cathode falls, and is due, as stated above, to excitation by electrons emerging from the Aston dark space. At higher cathode falls (loc. cit.) the cathode glow appears and becomes strong in Ne, N_2, H_2 and Xe, but remains weak in Ar, He and O_2 even with high cathode falls.

Further, electrostatic deflection experiments (see § 7) on the canal rays showed clearly that a large part of the luminosity is due to the presence of neutrals. Measurements indicate that the positive-ion stream to the cathode is not limited in extent to the region of luminosity, although the concentration of ions is larger at the centre of the tube (shown by Faraday-cylinder measurements and the burning of the cathode). Further evidence for the predominating influence of neutral particles in the formation of the pencil glow is produced when the relative importance of ionization by, and neutralization by electron transfer (*Umladung*) of, protons is considered.

These reactions may be represented as follows:—

(*a*) Ionization by protons: $H_1^+ + H_2 \rightarrow H_1^+ + H_1^+ + H_1^- + e^-$

or $H_1^+ + H_2 \rightarrow H_1^+ + H_2^+ + e^-$.

(*b*) *Umladung* (change of charge) of protons:

$$H_1^+ + H_1 \overset{\text{slow}}{\underset{\text{fast}}{\rightleftarrows}} H_1 + H_1^+.$$

The protonic mean free path at 0·15 mm. pressure is approximately, taking 7 kv. protons for which the Umladung effect is a maximum (Bartels, 1932), 6 cm. for ionization and 0·15 cm. for Umladung (Gerthsen, 1930). In these conditions, neutralization by electron transfer is more important than the production of more free electrons by a normal ionization process. It is difficult to assess the relative magnitudes of these two effects in a particular practical case. For instance, it is necessary first to determine the effective length of discharge path between anode and cathode in order that the p.d. for each mean free path may be found. This is difficult for the tubes described below (§ 6). If we assume an effective length of 3 cm. for 20 kv. discharge voltage, it is seen that the Umladung conditions given above (7 kv., m.f.p. = 0·15 cm.) are incompatible, although approximate corrections could be obtained by trial and error. It is hoped to discuss this problem further elsewhere.

The apparent diameter of the pencil glow in hydrogen in tubes with pressure $\sim$0·1 mm. Hg and $\sim$20 kv. operating potential for a current of $\sim$10 ma. is approximately 2 mm., a surprisingly small figure. In order to explain this fully it would be necessary to determine the secondary process responsible for the glow, i.e. excitation of one of the products of the change of charge reaction,

excitation or ionization by neutrals, recombination, etc. Bartels (1920) and Meyer (1937) stated that the cross-sections for ionization by neutral atomic hydrogen ions lie below the values given by Gerthsen (1930) for ionization by protons. See also the work of Yearian (1940).

The above discussion is not complete, but is intended to demonstrate the extreme importance of fast neutral particles in hydrogen glow discharges at pressures $\sim$0·1 mm. Hg.

Finally, the discharge mechanism as a whole should be outlined. (*a*) For high pressures ($\sim$1 mm. Hg) the discharge is maintained in stability by secondary emission from the cathode, the secondary electrons after acceleration producing a fresh supply of ions. (*b*) The pencil glow discharges can be maintained by ionization produced in the gas by the fast positive ions or neutrals.

Generally, therefore, for stability,

$$pn>1$$

where for case (*a*) n is the number of ion pairs produced in the gas by one electron, ignoring diffusion losses, and p the number of secondary electrons produced at the cathode per incident positive ion. For case (*b*) n has the same meaning, but p is now the number of fresh electrons produced by each positive ion in the gas (Seeman and Orban, 1935, and Hailer, 1938). It is clear that secondary emission at the cathode for case (*b*) plays only a small part because a large proportion of the total positive current arriving at the cathode may be collected after passing through the canal.

In order to estimate the efficiency of this type of tube as an ion source, it is necessary to determine the total positive current arriving at the cathode by experiment and to express this as a fraction of the discharge current, read by a milliammeter in series with the tube. Such calculations are only made with great difficulty (Druyvesteyn and Penning, 1940, p. 136). Campan (1934) found for an almost normal cathode fall that the electronic current at the cathode was 6 to 16% for H_2 and O_2, but his conditions were different from those obtaining in the pencil-glow régime. In our case secondary emission from the cathode would, presumably, only be of importance because of the presence of large numbers of fast neutral particles. A high proportion of the ions passed through the cathode, and from the geometry of the tube (§ 6) we assume that photo-emission was small.

§ 4. MEASUREMENT OF HIGH-SPEED POSITIVE-ION CURRENTS

If a flat plate is exposed to a beam of positive ions with an energy of several kv., the resulting emission of secondary electrons can be shown, by the action of a suitable accelerating field, to be large enough to vitiate measurements of the primary beam. The problem has been investigated by many workers, but we shall refer only to the experiments of Hill and collaborators (1939) on H_1^+, H_2^+ and He_4^+ ions. They used surfaces which had not been specially prepared, so the results, a selection of which is given in table 1, are of particular interest

with reference to continuously evacuated apparatus where baking is not practicable. For the following results, however, stable targets were obtained by continued degassing.

Table 1

Ion species	Target metal	Energy (kv.) 78	107	142	213	284	355	426
		Yield of secondaries/primary						
H_1^+	Cu	3·88	3·61	3·41	2·90	2·52	2·44	2·21
H_2^+	Cu	6·68	6·64	6·45	6·26	6·30	5·71	5·60

The results for Mo, Al and Pb were very similar, indicating that surface films were exerting a controlling influence. The secondary particles were very slow, e.g. for 213 kv. protons on Mo the great majority (~95 %) of secondaries could

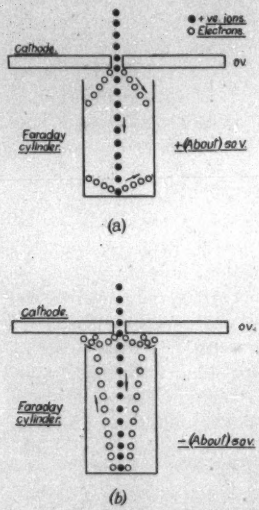


Figure 1. Effect of secondary emission of electrons on the measurement of positive-ion currents.

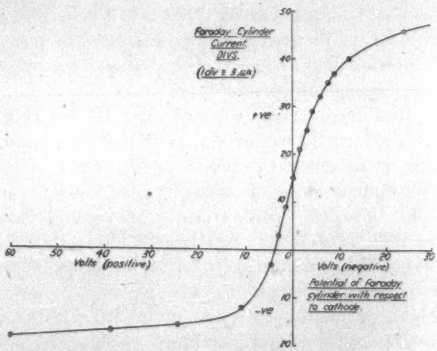


Figure 2. Effect of secondary emission of electrons on the measurement of positive-ion currents.

be turned back to the target by a retarding potential of 40 v. The increased emission with molecular ions is due to the latter splitting up into atomic ions (this effect is mentioned in § 2). In our case, an extra complication arises due to the proximity of the Faraday cylinder to the cathode (no focussing was used)

because of the secondaries arising in the canal. Consequently, a simple retarding field could not be used because the collector would receive negative particles from the canal. The effect is illustrated in figure 1. The characteristic of a simple Faraday cylinder, 7·5 cm. long × 2·5 cm. internal diameter, situated 3·7 cm. from the cathode, is shown in figure 2.

The problem was solved (see also Hailer, 1938) by the use of an intermediate electrode, maintained 120 v. negative with respect to the Faraday cylinder and 240 v. negative with respect to the cathode. The latter was, therefore, 120 v. positive with respect to the Faraday cylinder, but the smallness of the aperture in the intermediate electrode prevented field penetration. Emission of secondary electrons from the diaphragm into the Faraday cylinder because of bombardment by stray ions was very small. The final dimensions were:—Faraday (inner) cylinder 1 in. internal diameter, $^1/_{32}$ in. thick, 3 in. long; intermediate diaphragm $1^1/_2$ in. diameter, $^1/_{16}$ in. thick, aperture 1·1 cm. diameter spaced $^1/_8$ in. from the upper end of the inner cylinder, which, in turn, was spaced about 1 cm. from the cathode. The characteristics for this arrangement are shown in figure 3 ($\sim$10 kv. discharge). The outer cylinder served merely as a support for the diaphragm; both were at the same potential, and insulated from the inner cylinder.

Finally, mention should again be made of the neutral particles, which will not be affected by electric fields, and which may, in some cases, invalidate experiments such as those described above. We have ignored the existence of massive negative ions, which were later shown to be present only in very small numbers (§ 8). If it is desired to simplify ion tubes or similar devices in which collectors are required, other methods of eliminating spurious effects due to secondary emission are available, e.g. magnetic fields or the use of long narrow cylinders. In the latter case it may be taken as a rule that the field falls off to <5 % of its external value at a distance inside the collector equal to its diameter. These methods were used in the present research and found to be satisfactory. All the measurements given here, however, were taken with the composite cylinder.

§ 5. EXPERIMENTS WITH THE FIRST SOURCE

The methods of measurement outlined in § 4 were applied to a source of simple design (figure 4). The anode and cathode were water-cooled, the former being insulated from earth for 50 kv., with an isolated water system. The anode (B) and cathode were made of chromium-plated copper, to minimize field currents at the cathode and to give good thermal conduction. The flanges (C) were lapped and provided with centring screws (not shown). The cylinder (A), forming part of the cathode, was necessary to prevent waste discharges between C and the upper parts of B. The insulating cylinder was made of Pyrex, and the plate (D) of soda glass. Several cylinders (A) were made, each with a corresponding Pyrex cylinder, to give a constant {anode rim/cathode} spacing. Several lengths of anode chamber were tried.

Hydrogen was admitted through a small hole in the cathode (not shown in

figure 4) from a needle valve and reservoir. No difference was found between tank hydrogen (99·5 to 99·7 % hydrogen, most of the impurity being oxygen) and that diffused through a palladium leak. This is hardly surprising, in view of the sealing technique adopted, i.e. the use of bitumen or Apiezon Q compound joints. The source gas pressures were measured with G.E.C. Pirani gauges,

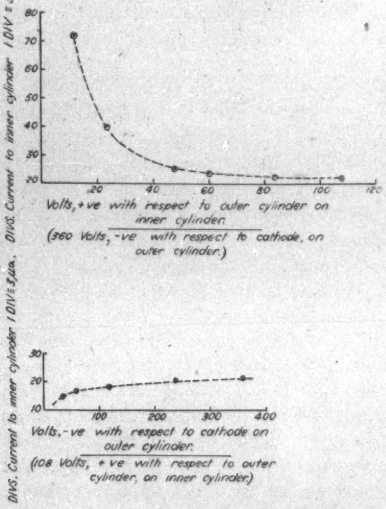


Figure 3. Experiments on composite Faraday cylinder.

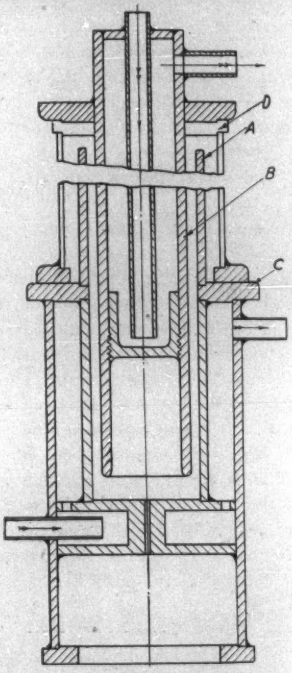


Figure 4. One-half size.

individually calibrated against a Macleod gauge and sealed on to a pipe soldered into the cathode block.

The pumping system (Metropolitan-Vickers oil-diffusion pumps) had a speed of about 20 l./sec. for air. The canal was 2 mm. diameter × 6·5 mm. long, these dimensions having been found by experiment to give the best yields for a reasonable gas flow through the tube.

A selection of results is given in figure 5 for an anode chamber 5 cm. long. The voltages ranged from 22 to 25 kv. for 0·092 mm. Hg. pressure and from 18·9 to 23 kv. for the higher pressure run.

Increasing the anode chamber to 10 cm. length gave slightly lower yields. Some results are given below:—

Table 2

5 cm. anode chamber				10 cm. anode chamber			
Hydrogen pressure (mm. Hg)	Tube voltage (kv.)	Tube current (ma.)	Canal-ray current (μa.)	Hydrogen pressure (mm. Hg)	Tube voltage (kv.)	Tube current (ma.)	Canal-ray current (μa.)
$0{\cdot}09_2$	22·2	2·1	60	$0{\cdot}08_8$	26·1	1	21
	23·4	2·8	75	,,	26·7	1·8	39
	24·2	3·6	90	,,	28	2·7	53
	24·8	4·4	105	,,	28·5	3·7	78
$0{\cdot}09_9$	16·2	1·6	40	0·11	13	1·2	25
	17·4	2·8	63	,,	13·4	1·8	33
	18·4	3·9	78	,,	13·5	2·3	40
	17·1	4·8	90	,,	13·6	2·8	42

A few readings at 15 cm. chamber length indicated that the yields were higher than those with a 10 cm. chamber and almost equal to those obtained with the 5 cm. chamber. Inserting a diaphragm with a 13 mm. central aperture in the lower end of the anode chamber increased the initial impedance, the required voltage for a given current and pressure being approximately doubled with reference to the data of figure 6—for instance, at 0·18 mm. Hg pressure 27·3 kv. (3·9 ma.) gave 141 μa. and 23 kv. (2·1 ma.) gave 96 μa.

The results may be briefly summarized. The yield increases steeply with main current but not with voltage. This indicates that the effective beam diameter does not increase rapidly with current. The cathode was usually burnt round the canal, indicating a beam diameter of $\sim$4 mm. Increasing the anode/cathode clearance from 2 mm. to 8·5 mm. made little difference to the performance of the tube—in fact it seemed that the beam diameter depended very little on electrode geometry for a tube without diaphragm, but no tests were made on very large tubes such as those used by Oliphant (1933), or Bouwers, Heyn and Kuntke (1937). The latter, working at pressures of 0·02 to 0·03 mm. Hg, obtained yields of 100 to 200 μa. at 25 kv. for discharge currents of 5 to 40 ma. respectively. Their data extend to 50 kv., 40 ma., at 0·021 mm. Hg pressure when the canal-ray yield was 600 μa.

It is of interest to mention that very small amounts of air mixed with the hydrogen caused the striking potential of the tube to be lowered considerably. This indicates that such sources might be operated at much lower voltages by the addition of an adulterant (e.g. nitrogen) without appreciably altering the proton content of the beam.

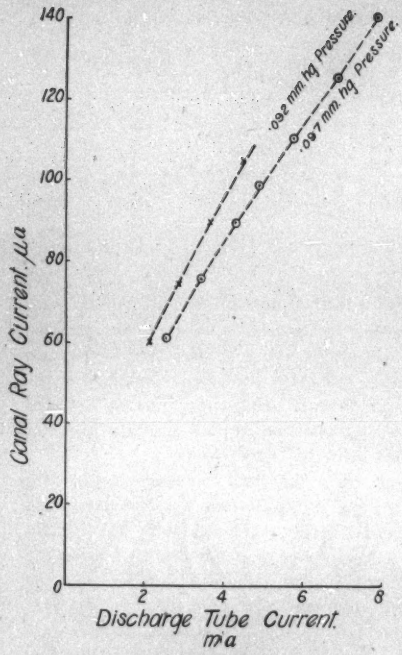


Figure 5.

Figure 5. Variation of canal-ray current with discharge-tube current (first tube).

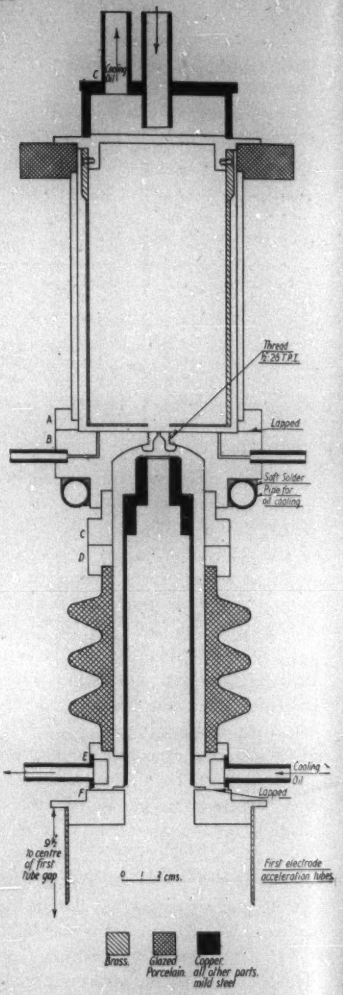


Figure 6.

Figure 6. The ion source and focussing electrode assembly. To avoid complicating the diagram, the following details are omitted: (1) Bakelite clamping bolts and nuts for ion source with brackets on B and C; (2) centring screws bearing on A with brackets on B; (3) centring screws bearing on C with brackets on D; (4) bakelite clamping bolts with brackets on B and F; (5) centring screws bearing on E with brackets on F; (6) steel clamping bolts with brackets on F and first plate of acceleration tube. Interior surfaces of ion source are polished.

§ 6. EXPERIMENTS WITH THE FINAL SOURCE

Many variations of the first source were tried, and simplifications resulted, a representative design being shown, with focussing assembly,* in figure 6. The tube diameter was increased in the hope that the operating pressure would be thereby reduced. The cathode cylinder was finally mounted on a lapped flange for screw centring (as in figure 4), the tube of figure 6 being centred

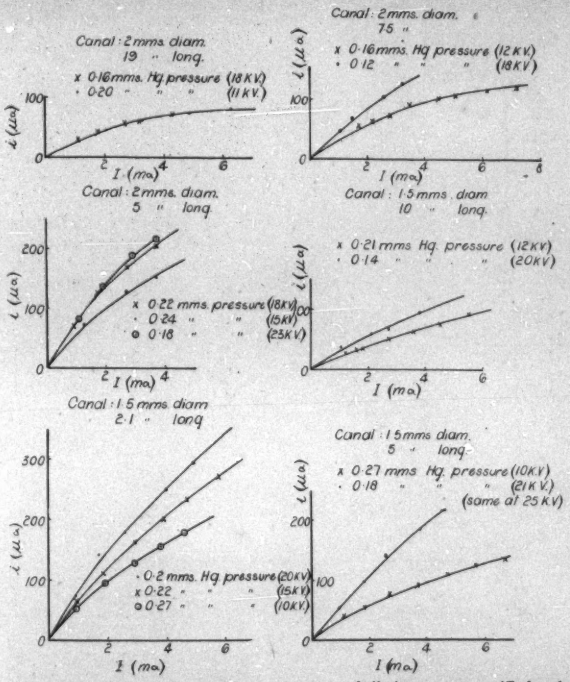

Canal: 2 mms. diam.
19 " long.
x 0·16 mms. Hg pressure (18 KV)
· 0·20 " " " (11 KV.)
Canal: 2 mms. diam.
7·5 " long.
x 0·16 mms. Hg pressure (12 KV)
· 0·12 " " " (18 KV)
Canal: 2 mms. diam.
5 " long.
x 0·22 mms. pressure (18 KV)
· 0·24 " " (15 KV)
o 0·18 " " (25 KV)
Canal: 1·5 mms. diam.
10 " long.
x 0·21 mms. Hg. pressure (12 KV)
· 0·14 " " " (20 KV)
Canal: 1·5 mms. diam.
2·1 " long
· 0·2 mms. Hg pressure (20 KV)
x 0·22 " " " (15 KV)
o 0·27 " " " (10 KV)
Canal: 1·5 mms. diam.
5 " long.
x 0·27 mms. Hg. pressure (10 KV)
· 0·18 " " " (21 KV.)
(same at 25 KV)
i (μa)
I (ma)

Figure 7. Canal-ray currents (i) plotted as a function of discharge current (I) for the type of tube shown in figure 6. Faraday cylinder 2·5 cm. from exit end of canal.

mechanically. The anode insulation consisted of a porcelain ring (external diameter 6 in., thickness $\frac{3}{4}$ in., internal diameter $3\frac{1}{2}$ in.), a ground recess being provided for the cathode cylinder. The tube length was 6 in., diameter of anode 3 in. The anode diaphragm aperture, the only critical dimension, was finally 13 mm. diameter, spaced 3 mm. from the cathode. The anode cathode side clearance was about 5 mm. With minor alterations, figure 6 represents the final design of tube.

* The problem of focussing ion-beams will be the subject of a further paper.

It will be seen that the construction of canal-ray sources presents no mechanical difficulties; the centring is critical, but lining up is practicable mechanically.

The rectifier set (50 kv. maximum) for supplying power to the tube was provided with a Philips Metalix valve (Type 28202) for experiments with currents up to 30 ma.

It was necessary to age the tube for periods of almost an hour for stability at currents above 5 ma., but even prolonged treatment with many different

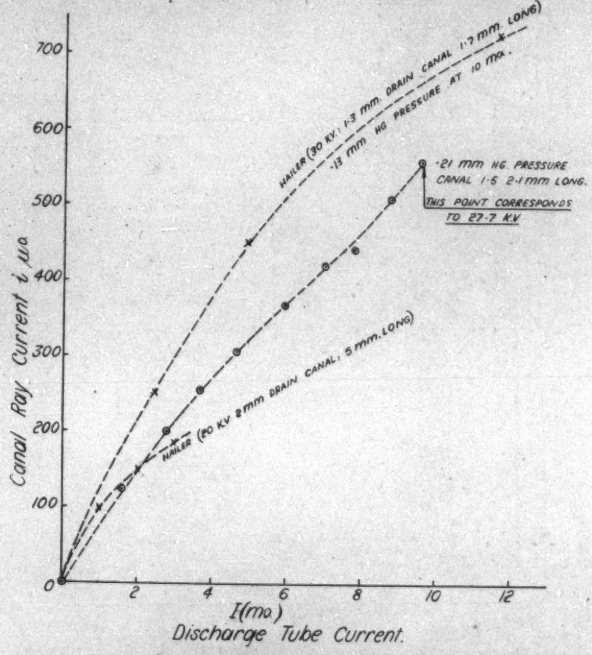


Figure 8. Variation of canal-ray current with discharge-tube current (second discharge tube).

forms of anode/cathode insulation showed that 35 kv. was the technical upper limit of voltage for our tubes. Attempts to increase the internal spark-over distances invariably lead to the establishment of stray discharges.

Preliminary tests of this source with low currents are shown in figure 7. The purpose of this work was to establish the most favourable form of canal. It seemed that, although the gas consumption was high, we should concentrate on the 2 mm. diameter × 5 mm. long and 1·5 × 2·1 mm. long (approximately

same pumping speed) canals. Figure 8 shows a more detailed curve for the latter, and includes some of Hailer's results (1938). Figures 9 *a* and *b* extend the data to higher currents, with the Faraday cylinder 2·5 cm. from the canal exit. Finally, figures 9 *c*, *d* and *e*, with a {cylinder/cathode exit} clearance of 1·2 cm., give the highest yields so far obtained with our tube. The differences in curvature between figures 9 *a* and *b* and figures 9 *c*, *d* and *e* were confirmed several

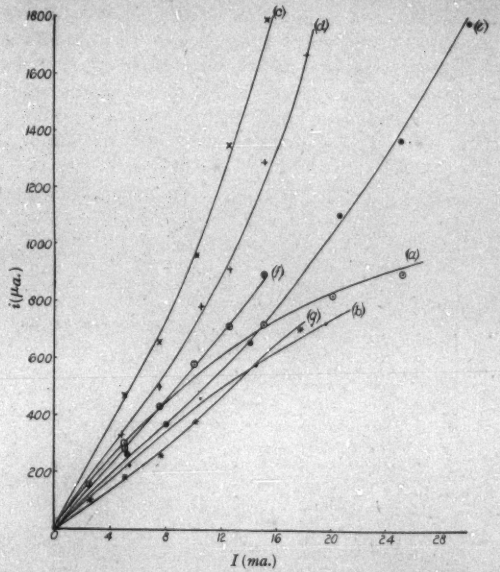


Figure 9. • Canal 2 mm. diam. : 5 mm. long. 0·16 mm. Hg. pressure (25 kv.).
⊙ „ 1·5 „ : 2·1 „ 0·15$_5$ „ „ (26 kv.).
× „ „ „ : „ „ 0·15 „ „
\+ „ „ „ : „ „ 0·17 „ „
● „ „ „ : „ „ 0·18 „ „
⊕ Nitrogen : canal 1·5 mm. diam. : 2·1 mm. long. 0·06 mm. Hg pressure 25 kv.
✳ „ : „ „ „ : „ „ 0·06$_5$ „ „ 15 kv.

times. It is hoped to give more data on this subject in a later publication. Figure 9 was taken with a 13-mm. diameter aperture in the anode diaphragm. In order to give as high a proportion of atomic ions as possible, the importance of which has been stressed above, a short canal is not beneficial, because the pressure gradient is high and the ion beam has little chance of losing intensity because of the ionization and Umladung processes. The primary particles are H_2^+; hence the probability of dissociation leading to the formation of H_1^+

ions is small. On the other hand, with a long canal, the beam emerging from the discharge will traverse in the canal a longer path in gas at higher pressures, so that the opportunity for dissociation will increase but the emergent-beam intensity will decrease. A broad and long canal should, therefore, give the greatest yield with a low proton content (Schütze, 1939). For a given beam-strength the gas flow should always be arranged to be a minimum.

The fractional yield in general falls off at higher currents, presumably because of space-charge effects in the source (see, however, figures 9 *c*, *d* and *e*, which are unusual in this respect). It is believed that if the voltage were raised to 50 kv. the yields at ~20 ma. would be much increased.

It seems beneficial in many respects to take curves such as figure 9 at constant pressure and varying voltages rather than at constant voltage and varying pressure. It is always advisable to plot the current as one of the variables.

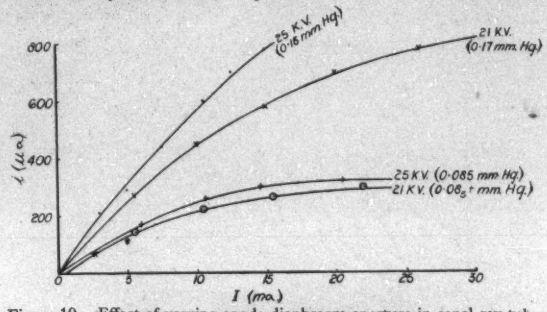


Figure 10. Effect of varying anode diaphragm aperture in canal-ray tube. Voltages and pressures given on curves.

13 mm. aperture in anode diaphragm: • 25 kv. × 21 kv.
25·4 mm. „ „ „ + 25 kv. ⊙ 21 kv.
No diaphragm ✻ 30 kv.

The principal reason for the choice of plotting i/I was that the Pirani gauges, despite constant tapping, were prone to give variable readings, due largely to zero shifts (contact resistances, etc.). It was found that for a given current at a given voltage the yield was approximately repeatable from day to day even though the Pirani readings showed considerable variations. This was not considered a serious disadvantage, as the yield figures were meant to be relative only, provided that the pressures were known approximately, as the gas consumption of the sources is an important quantity. The latter is, unfortunately, high for tubes of the cold-cathode type with diaphragm.

Removal of the focussing diaphragm gave lower striking voltages. For instance, at 0·05 mm. Hg. hydrogen, 10·5 kv. sufficed to give 2 ma. (cf. figure 7). The yields obtained were low, e.g. 6 ma., 17 kv., 55 μa. and 5·5 ma., 33 kv.,

115 μa. (pressures $\sim$0·05 mm. Hg). These figures agree fairly well with those of Bouwers, Heyn and Kuntke (1937), also obtained without focussing diaphragm. Figure 10 shows the effect of increasing the size of the aperture in the diaphragm in our tube, and includes two points taken without a diaphragm. The reduction in striking voltage for increasing aperture size is to be noted. For all these tests the {anode diaphragm/cathode} clearance was about 4 mm.; the Faraday cylinder for figure 10 was 1·0 cm. from the exit end of the canal.

Voltage/current curves for the tube are shown in figure 11. Some yield-figures for tank nitrogen ($\sim$99·5 % pure) are added (figure 9) to show the differences in striking potentials (effect of increased efficiency of ionization, see

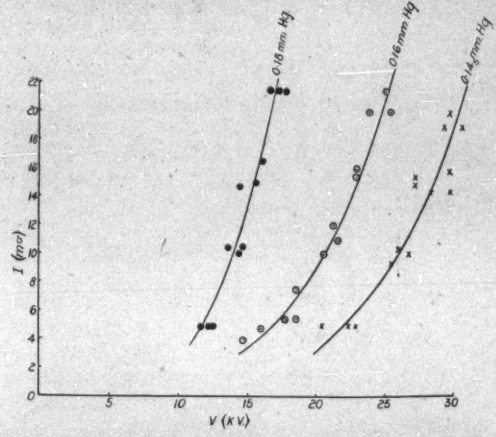


Figure 11. Current/voltage relation for canal-ray tubes 13 mm. anode diaphragm. Pressures (hydrogen) marked on curves.

Zwicky (1929)). The improved performance with small-diaphragm apertures may be due to increased pressure effects (" gas focussing ") or better electron-optical focussing or both. The anode diaphragm and cathode may be considered (Myers, 1939) as a pinhole lens, the focal length (f) of which may be expressed as

$$\frac{1}{f} = \frac{V_2' - V_1'}{4V_g},$$

where V_g is the axial potential at the diaphragm, and V_2' and V_1' the potential gradients in the image and object space respectively. It should be noted that such lenses are not comparable with optical pinhole lenses, which do not have sharply defined focal lengths. Put $V_1' = 0$, $V_2' = \dfrac{20 \times 10^3}{0{\cdot}4}$ v./cm., since the {anode diaphragm/cathode} clearance is 0·4 cm. and 20 kv. a representative tube voltage.

Since the treatment is clearly approximate, assume an infinite object distance (in fact ~10 to 20 cm.), so that f gives the image distance, which is required to be about 0·4 cm.; hence the beam is focussed on to the canal. V_0 is thus found to be about 5 kv., which seems reasonable according to the published data on field plots. Our pinhole, of course, is of finite size, and errors are thereby introduced. However, we may assume that the focussing in our tubes at about 20 kv. is satisfactory, but the probability of a "cross-over" in the canal with subsequent beam divergence should not be overlooked. This is, perhaps, the most serious defect of sources of this type, as post-cathode focussing is rendered more difficult. It must be emphasized that a high beam current is required usually at the target of some high-voltage tube, probably at 6 to 12 ft. from the canal. Consequently, to obtain a high-current canal ray is only part of a formidable problem.

Finally, we should discuss canal sizes, and their bearing on tube yields, from the point of view of gas consumption and acceleration tube pressure. Consider the latter: most workers quote tube pressures of 2 to 5×10^{-5} mm. Hg., but Bouwers, Heyn and Kuntke (1937) operated a tube satisfactorily at 8×10^{-4} mm. Hg. Parkinson *et al.* (1938) stated that defocussing took place if the tube pressure exceeded $1{\cdot}5\times10^{-5}$ mm. Hg., in conflict with the findings of most workers. Assuming an acceleration-tube pumping speed of 300 l./sec., which is easily obtainable, an acceleration-tube pressure of 5×10^{-4} mm. Hg., and a source pressure of 0·2 mm. Hg., the canal pumping-speed must not exceed 750 c.c./sec. When using expensive gases, e.g. deuterium, it is, of course, necessary to reduce gas consumption as much as possible. The pumping system could be modified to circulate the gas, but this is generally not considered practicable. Using Knudsen's formula (Knudsen, 1909) for the pumping speed of narrow tubes (length $\gg$ radius; mean free path $\ll$ geometrical dimensions) we obtain 380 c.c./sec. for a 1·5 mm. diameter canal 2·1 mm. long, so that a reasonable safety factor is obtained. From a more accurate formula (Clausing, 1932), which allows for dimensions ~ mean free paths, the calculated speed was about 350 c.c./sec. (within ±5%). We may take the pumping speed S approximately as

$$S=k_1D^3/l,$$

where D = canal diameter, l = canal length, and k_1 is a constant. The output current varies approximately as $1/l$ (figure 7) until very short lengths are reached, and is independent of D, provided the canal area is greater than the focal spot, or the area (a) of the beam at the cathode. If $\pi D^2/4 \ll a$, the beam will be approximately proportional to D^2 for constant beam-current density. In practice it is found that $D=1{\cdot}5$ mm. is sufficiently large (see figure 7) for the beam current (i) to be given approximately by $i=k_2/l$ for varying D and l. Hence in general it is better to decrease D in order to reduce the pumping speed. This tends (see above) to increase the proton content of the beam.

§ 7. ELECTROSTATIC DEFLECTION OF CANAL RAYS

In the absence of detailed knowledge of ionization processes in the tube it it was considered worth while to investigate the possibility of electrostatic separation of the ion species. It is known, however, that for a given charge, the small angular deflection, θ, suffered by a charged particle (mass m, velocity v) travelling through an electrostatic field is proportional to

$$1/mv^2.$$

Consequently, if all ions (H_3^+, H_2^+, H_1^+) were to be formed at the same point in the discharge space, i.e. fall through the same potential difference, mv^2 would be a constant and there would be no separation. It was not known with certainty that this mechanism would obtain, so in view of other useful information that could be gained, a few tests were made.

Two deflector plates (5 cm. long × 6 cm. wide, with a separation of 2 cm.) were mounted in a horizontal pump-manifold 1 cm. behind the canal and observed through a glass plate in the end of the chamber.

It was found that perhaps 70 to 80 % of the luminosity (" hydrogen red ") of the canal rays was due to neutral particles, as the application of even 50 kv. failed to give detectable deflection of the main beam. The faint deflected beam was moved through 4·5 cm. at 25 cm. from the centre of the deflector system for 640 v. applied to the plates (beam positive), and gave a faint fluorescent patch on the glass plate. There was no visible separation of ion species; this was confirmed by Faraday-cylinder measurements. About 15 % of the no-field current was obtained when a large deflecting potential was applied to the plates, due possibly to liberation of some negative secondaries from the collector struck by fast neutral hydrogen atoms, but probably to neutrals gaining a positive charge by a reverse Umladung process after leaving the deflecting field.

The only point of interest in this experiment was the large proportion of neutral particles found in the emergent canal-ray beam. The estimate was, of course, entirely visual, and only has meaning if we assume the neutral and charged beams to be equally efficient in causing emission of light.

§ 8. BEAM ANALYSIS BY ELECTROMAGNETIC DEFLECTION

The importance of a high proportion of atomic ions in the beam has been stressed in § 2. Analysis by electromagnetic deflection only is sufficient (momentum resolution) if the velocity range of each group of particles at the target of the analyser is not high.

Some photographs of the spectrum obtained by allowing the analysed beam to strike a fluorescent screen were taken. It was not possible to perform tests on the ion source alone, owing to lack of time, and all the results given here refer to a high-voltage (~400 kv.) beam analysed by an electromagnet and suitable slit system at the earthed end of an acceleration tube.

The slit width was 3 to 4×10^{-3} in., $\frac{1}{2}$ in. long, cut in a sheet of copper 10×10^{-3} in. thick. The slit was mounted in a flange at the entrance to a Pyrex

tube, about 6 cm. diameter, which passed between, and extended 12 cm. beyond, the 4-in. pole pieces. The total distance from screen to slit was 23 cm. A beam current of $\sim$1 μa. was sufficient to give reasonable contrast, higher currents increasing the background illumination too much for satisfactory measurement.

The proton line was identified first by measurement, and the identity of other lines deduced from a knowledge of their deflections (see table 3) taken from the fluorescent screen.

H_{2-1}^+ ions are those which enter the tube as H_2^+ ions and which split up in the acceleration tubes (eight gaps, so that eight ill-defined lines would be expected in the region 6–7 of table 3). If the field gradient were constant along the tube, and not concentrated at the focussing gaps, a continuous range of energies for the H_{2-1}^+ ions should be obtained. Some of the lines in table 3 are not visible in the photographs, so sketches of the spectrum are given in figure 12.

Table 3

Line number	Assumed identity of ions	Deflections (cm.)	
		Calculated	Measured
0	neutrals	0	0
1	H_3^+	1·65	1·65
2	H_2^+	2·0	2·0
3	H_2 or H_3	—	2·15
4	H_2 or H_3	—	2·3
5	H_{2-2}^+	2·45	2·55
6	H_1^+	2·85 assumed	2·85
6–7	H_1^+ and H_{2-1}^+	—	—

The lines 3 and 4 were due to molecular ions formed in the tube; as the upper (high-potential) stages of the tube were shorted out, keeping the total voltage constant, and thus increasing the field-free region at the top of the tube, lines 3 and 4 merged into line 2. As the lower gaps were successively connected to earth, the lines 3 and 4 moved into line 6. For example, if a proton falls through half the accelerating-tube potential-difference before losing its charge to a neutral molecule (H_2) or attaching itself to an un-ionized atom, then the deflection for a given deflecting field will be $[\sqrt{(\frac{2}{3})} \times$ proton deflection$]$. If this explanation of lines 3 and 4 is correct, it is surprising that there are only two or three lines in the spectrum for these processes. It is possible that they were due respectively to H_3^+ and H_2^+ (both formed in the accelerating tube). In the absence of more exact data, no definite conclusions may be drawn, but the reversed Umladung process should not be excluded as a possibility.

As mentioned above, we were not able to perform beam analysis on the ion source without acceleration tube. The heavier particles would probably be scattered more in the latter by space-charge effects than the atomic ions, though

the current density, and, therefore, radial forces, were small in these experiments. In general, also, the effective absorption (due to all the collision processes) will be appreciably more for the molecular ions. The accelerating tube was operated at a pressure of $\sim 10^{-4}$ mm. Hg., and was 8 ft. long. In order to gain information on the beam whilst minimizing the complicating effects of the tube, a few tests were made with one focussing gap (35 kv.) only in operation, the other electrodes being connected to earth. The spectra were too faint to be satisfactorily photographed, but taking the brightness of the H_1^+ line as 10, that of the H_2^+, H_3^+,

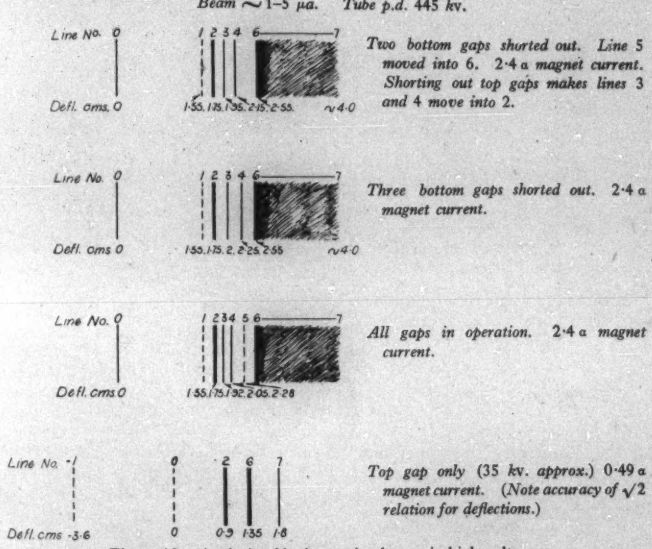

x displacements to scale, ×2 Pole separation 6·5 cm. Source ~ 9 kv. ~ 0·5 ma.
Beam ~ 1–5 μa. Tube p.d. 445 kv.
Line No. 0
Defl. cms. 0
1 2 3 4 6 7
1·55 1·75 1·95 2·15 2·55 ~4·0
Two bottom gaps shorted out. Line 5 moved into 6. 2·4 α magnet current. Shorting out top gaps makes lines 3 and 4 move into 2.
Line No. 0
Defl. cms 0
1 2 3 4 6 7
1·55 1·75 2 2·25 2·55 ~4·0
Three bottom gaps shorted out. 2·4 α magnet current.
Line No. 0
Defl. cms 0
1 2 3 4 5 6 7
1·55 1·75 1·92 2·05 2·28
All gaps in operation. 2·4 α magnet current.
Line No. -1
Defl. cms -3·6
0 2 6 7
0 0·9 1·35 1·8
Top gap only (35 kv. approx.) 0·49 α magnet current. (Note accuracy of √2 relation for deflections.)

Figure 12. Analysis of hydrogen-ion beams in high-voltage acceleration tube.

and neutral lines respectively were estimated visually as 7–8, 2–3 and 3–4. Some negative ions were present in very small numbers; for the high-voltage work none were observed.

Some quantitative measurements were taken with a Faraday cylinder, using larger deflections obtained with a special assembly (35° deflection, so that even with a $\frac{1}{4}$-in. slit width the resolution was adequate), figure 13 showing a typical curve. The calculated magnet field current for proton detection was 3·2 amp.; the $\sqrt{2}$ relation between the currents for H_2^+ and H_1^+ usually held only approximately, because the stray fields for this experiment were larger than for the

photographic work described above, where the $\sqrt{2}$ factor was accurately obtained. The proton content was found from the curves (e.g. figure 13) to be about 40

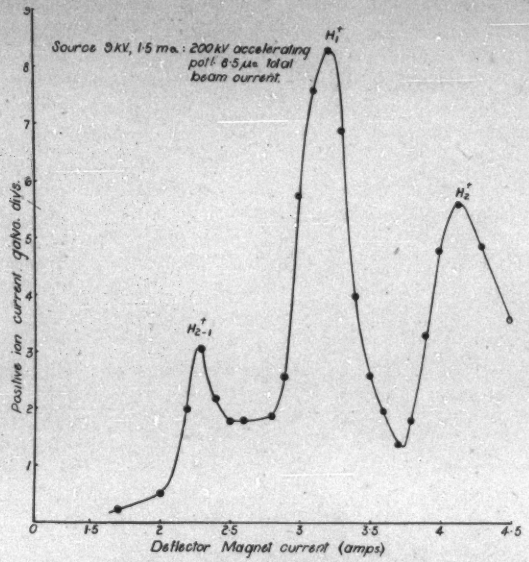


Figure 13. Analysis of hydrogen-ion beams in high-voltage acceleration tube (*continued*).

to 50 %. No appreciable dependence of beam composition on source conditions was found, but it should be pointed out that the source only operates for a limited range of pressures, i.e. 0·08 to 0·25 mm. Hg.

§ 9. CONCLUSIONS

A detailed study of the behaviour of canal-ray positive-ion sources for use with atomic disintegration apparatus gave the following results: (*a*) the total yield of positive ions could be made high (1 to 2 ma.); (*b*) the efficiency is low, though this usually is not of any importance; (*c*) the gas consumption is in general high; (*d*) the proportion of atomic ions is satisfactory, i.e. about 40 to 50 %.

This paper does not deal with the problem of focussing the canal-ray beam; given focussing voltages ~50 kv., this can be satisfactorily carried out. It is hoped to contribute another paper on this subject shortly.

§ 10. ACKNOWLEDGEMENTS

This research was carried out as part of a programme, on a high-voltage apparatus for atomic disintegration experiments, initiated and directed by Professor C. D. Ellis, F.R.S., whom the author wishes to thank for guidance and encouragement.

The author also wishes to express appreciation of the interest shown by Dr. T. E. Allibone, whose suggestions and advice have considerably helped the work.

Thanks are due to the Department of Scientific and Industrial Research for a grant covering part of the time during which the experiments were carried out, and to Dr. A. P. M. Fleming, Director and Manager, Research and Education Departments, Metropolitan-Vickers Electrical Co., Ltd., for permission to publish this paper.

REFERENCES

Amaldi, E., Bocciarelli, D., Rasetti, F. and Trabacchi, G. C., 1939. *Ricerca Scientifica*, **10**, 623.
Bartels, H., 1930. *Ann. Phys. Lpz.*, **6**, 957.
Bartels, H., 1932. *Ann. Phys., Lpz.*, **13**, 373.
Bothe, W. and Gentner, W., 1937. *Z. Phys.* **104**, 685.
Bothe, W. and Gentner, W., 1939. *Z. Phys.* **112**, 45.
Bouwers, A., Heyn, F. A. and Kuntke, A., 1937. *Physica*, **4**, 153.
Campan, T. I., 1934. *Z. Phys.* **91**, 111.
Chang, W. Y., Goldhaber, M. and Sagane, R., 1937. *Nature, Lond.*, **139**, 962.
Clausing, P., 1932. *Ann. Phys., Lpz.*, **12**, 961.
Druyvesteyn, M. J. and Penning, F. M., 1940. *Rev. Mod. Phys.* **12**, 87.
Franzini, T., 1938. *Atti Accad. Lincei*, **28**, 292.
Gerthsen, C., 1930. *Ann. Phys., Lpz.*, **5**, 657.
Günterschulze, A. and Keller, F., 1931. *Z. Phys.* **72**, 143.
Hailer, C., 1938. *Wiss. Ver. Siemens-Werken*, **17**, 321.
Hill, A. G., Buechner, W. W., Clark, J. S. and Fisk, J. B., 1939. *Phys. Rev.* **55**, 463.
Holst, G. and Oosterhuis, E., 1923. *Phil. Mag.* **46**, 1117.
Klemperer, O., 1939. *Electron Optics* (Cambridge : The University Press), p. 19.
Knudsen, M., 1909. *Ann. Phys., Lpz.*, **28**, 75.
Lamar, E. S., Buechner, W. N. and Compton, K. T., 1937. *Phys. Rev.* **51**, 936.
Lamar, E. S., Samson, E. W. and Compton, K. T., 1935. *Phys. Rev.* **48**, 886.
Laporte, M., 1939. *Décharge Electrique dans les Gaz* (Paris : Colin), p. 156.
Meyer, H., 1937. *Ann. Phys., Lpz.*, **30**, 635.
Myers, L. M., 1939. *Electron Optics* (London : Chapman and Hall), p. 166.
Oliphant, M. L. E., 1935. Int. Conf. Physics, London, 1934. *Rep. Progr. Phys.* **1**, 145.
Oliphant, M. L. E. and Rutherford, E., 1933. *Proc. Roy. Soc.* A, **141**, 259.
Parkinson, D. B., Herb, R. G., Bernet, E. J. and McKibben, J. L., 1938. *Phys. Rev.* **53**, 642.
Roberts, R. B., 1937. *Phys. Rev.* **51**, 810.
Schütze, W., 1938. *Wiss. Ver. Siemens-Werken*, **17**, 341.
Seemann, H. and Orban, G., 1935. *Ann. Phys., Lpz.*, **23**, 137.
Timoshenko, G., 1938. *Rev. Sci. Instrum.* **9**, 187.
Yearian, H. J., 1940. *J. Chem. Phys.* **8**, 24.
Zinn, W. H., 1937. *Phys. Rev.* **52**, 655.
Zwicky, F., 1928. *Proc. Nat. Acad. Sci., Wash.*, **18**, 314.

THE DETERMINATION OF THE OPTICAL PROPERTIES OF THICK MAGNETIC LENSES, AND THE APPLICATION OF THESE LENSES TO BETA-RAY SPECTROMETRY

By R. E. SIDAY, B.Sc.,
Birkbeck College, London

Communicated by Prof. J. D. Bernal, F.R.S. Received 24 November 1941

ABSTRACT. An examination is made of the validity of the usual mathematical expressions from which the focusing properties of axially symmetrical magnetic fields are deduced, and conditions are established for good image formation.

A method of ray tracing under these conditions is developed and applied to a particular thick lens consisting of a solenoid wound with multiple layers and shielded with soft iron. The focal lengths and the positions of the principal planes are determined in terms of a parameter f, which is the focal length of a thin magnetic lens having the same value of $\int H^2 dx$. The dependence of the true focal length and the separation of the principal planes on f is obtained in a generalized form applicable to the determination of these quantities for any thick lens wound as a solenoid with multiple layers.

A differential method of applying the above conditions for good image production is used for an examination of the merits of coils with different types of windings, and for the estimation of the errors arising for any particular coil.

The advantages derived in β-ray spectrometry from the use of thick magnetic lenses are discussed, and a method is proposed for increasing the resolving power of the single-lens spectrometer in which use is made of a large rotation of the line image formed of a line object which can be obtained by winding a lens of suitable length.

§1. THE THEORY OF THE MAGNETIC LENS

The simple theory of the magnetic lens (see, for example, Myers' *Electron Optics*) assumes that an electron at any point in an axially symmetrical magnetic field experiences a rotation about the axis whose velocity is

$$\frac{d\phi}{dt} = \frac{1}{2} \cdot \frac{e}{m} \cdot H_x, \qquad \ldots\ldots(1)$$

where H_x is the axial component of the magnetic field at the point considered, and that this rotational velocity produces an acceleration towards the axis which causes the focusing action of the lens. The usual cylindrical co-ordinates r, ϕ and x are used. The image is oriented at an angle ϕ to the object, where

$$\phi = \frac{1}{2} \frac{e}{m} \int \frac{H_x}{v_x} dx, \qquad \ldots\ldots(2)$$

v_x being the axial component of the velocity of the electron, and the integration being taken along each ray.

It is convenient (and legitimate) to consider the focusing of the lens and the rotation of the image separately, in which case the focusing of the lens becomes strictly analogous to that of an optical lens.

If v_x is assumed to be constant, equation (1) with the other equation of motion leads to

$$\frac{d^2r}{dx^2} = -\frac{r}{4}\left(\frac{e}{m}\right)^2 H_x^2/v_x^2, \qquad \ldots\ldots(3)$$

which is the equation of the trajectory of the electron within the lens. For a thin lens Busch (1926 and 1927) calculated the focal length,

$$\frac{1}{f} = \frac{1}{4}\left(\frac{e}{m}\right)^2 \int_{-\infty}^{+\infty} \frac{H_x^2}{v_x^2} dx, \qquad \ldots\ldots(4)$$

by directly integrating equation (3), assuming that r was constant within the lens for any electron. The constancy of r is only true in general for a thin lens.

These expressions show, however, how to obtain the condition for the accurate focusing of any magnetic lens. In this paper, only rays that are not skew to the axis outside the lens are considered. Equation (1) is accurately true everywhere in an axially symmetrical field if the equation

$$\frac{1}{2} \cdot \frac{\partial H_x}{\partial x} = -\frac{H}{r} \qquad \ldots\ldots(a)$$

holds at all points of the field. Except in the case of an infinite uniform field, this condition is not strictly consistent with Laplace's equation, although at a point in an axially symmetrical field it is more nearly satisfied the closer this point is to the axis.

Equation (2) shows that $\int \frac{H_x}{v_x} dx$ must be the same for all rays if the image of an object is to be oriented with respect to the object without angular spread. Now, in order to obtain equation (3) from equation (1), it is necessary to assume that the axial component of the velocity of the electron is unchanged in its passage through the lens. This leads to the condition

$$v_x = \text{constant}, \qquad \ldots\ldots(b)$$

while the condition of formation of images oriented without angular spread therefore reduces to

$$\int H_x dx = \text{constant}, \qquad \ldots\ldots(c)$$

where the integration is taken along the trajectory.

The degree of accuracy with which the lens produces equal rotation for all beams depends on the degree of accuracy with which the above conditions (*a*), (*b*) and (*c*) can be satisfied at all points along the rays.

Identical total rotation for all rays originating from the same point is a

condition for accurate focusing. Another condition follows under the same assumptions from (4), by integrating along the trajectory:—

$$\int H_x^2\,dx = \text{constant}, \qquad \ldots\ldots(d)$$

a condition independent, in general, of the constancy of $\int H_x\,dx$.

In geometrical optics, the defects of the image are exactly and elaborately classified, and this classification is of very great importance in the construction of lenses free from particular aberrations. It is not clear that such a procedure has the same importance for the magnetic lens. There is no method of combining lenses to reduce a particular aberration. In the design of a lens, emphasis should be placed on direct construction to the required standard of focusing rather than on the measurements or calculations of aberrations with a view to their subsequent correction. This can be effected by satisfying the four conditions above.

§ 2. THE THICK MAGNETIC LENS

Equation (3) shows that the *thin* magnetic lens will only focus to the degree of accuracy with which the above conditions are satisfied, and that equation (4) gives the focal length of the lens.

If the magnetic field spreads along the axis for a distance that is not small compared with the focal length, the lens is thick, r is no longer even approximately constant for an electron during its passage through the lens, so that while (2) still gives the rotation, and (3) still determines the paths of the rays when the above conditions hold, the focal length will no longer be given by (4).

In order to examine the rules for the focusing of a thick lens, such a lens was constructed as a solenoid about 20 cm. long, wound with multiple layers. The field along the axis was measured, and assuming that the above conditions were satisfied, the trajectories of electrons were traced through the lens, in order to determine its focal length, principal planes and its focusing properties.

§ 3. RAY TRACING THROUGH A THICK LENS

Rays were traced in terms of the parameter f defined by equation (4) above. In order to make the equation depend only on relative values of H, and to introduce f explicitly, equation (3) is written in the form

$$\frac{d^2r}{dx^2} = -\lambda . r\,kH_x^2, \qquad \ldots\ldots(5)$$

where k is an arbitrary constant arising from the measurement of H. So that now

$$\left(\frac{dr}{dx}\right)_2 - \left(\frac{dr}{dx}\right)_1 = -\frac{(r_1+r_2)}{2}\,.\lambda.\int kH_x^2\,dx, \qquad \ldots\ldots(6)$$

and if between two points on the axis of the lens

$$\int_1^2 kH_x^2\,dx = {}_2A_1, \qquad \ldots\ldots(7)$$

$$\left(\frac{dr}{dx}\right)_2 - \left(\frac{dr}{dx}\right)_1 = -\frac{(r_1+r_2)}{2}\cdot\frac{{}_2A_1}{fA}, \qquad \ldots\ldots(8)$$

where

$$A = \int_{-\infty}^{\infty} kH_x^2 dx. \qquad \ldots\ldots(9)$$

For purposes of ray tracing, the region over which the field extended was divided into 1 cm. steps, so that

$$r_2 = r_1 + \frac{1}{2}\left\{\left(\frac{\partial r}{\partial x}\right)_2 + \left(\frac{dr}{dx}\right)_1\right\}, \qquad \ldots\ldots(10)$$

$$\left(\frac{dr}{dx}\right)_1\left(1-\frac{{}_2A_1}{4fA}\right) - \left(\frac{\partial r}{\partial x}\right)_2\left(1+\frac{{}_2A_1}{4fA}\right) = r_1\cdot\frac{{}_2A_1}{fA}. \qquad \ldots\ldots(11)$$

Thus, knowing an initial r, and dr/dx, the rest of the ray can be traced. These initial points are taken outside the magnetic field. The f's appearing in these

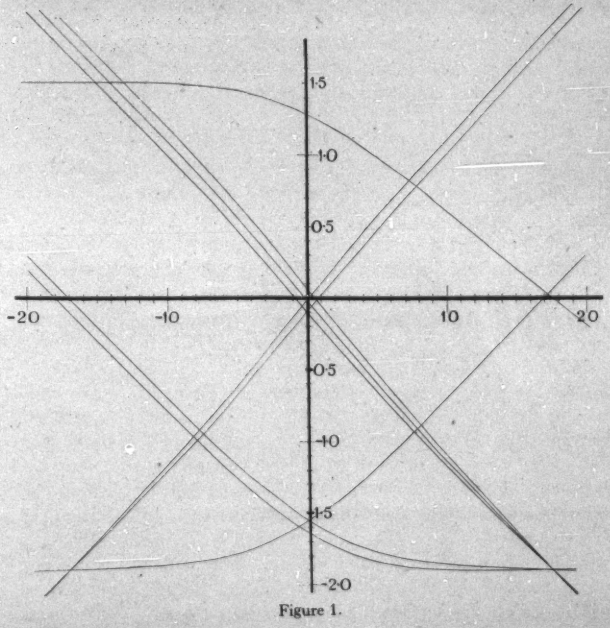


Figure 1.

formulae are the focal lengths the lens would have if it were infinitely thin, and they can be directly chosen in cm. for purposes of ray tracing. ${}_2A_1/A$ for any interval will depend only on the geometrical form of the winding. By tracing suitable rays, the real focal lengths, and the positions of the principal planes corresponding to different values of f, were determined.

This method of ray tracing gives the value of r at any point on the ray more accurately than it gives dr/dx at this point. The cardinal planes were therefore fixed by tracing a parallel bundle of rays in the object space inclined at some definite angle to the axis. These intersect in the image space on the image focal plane, and the ordinate of the intersection is related to the object focal length by the relation

$$r = F \,.\, \tan\theta, \qquad \ldots\ldots(12)$$

where F is the true focal length of the lens and θ the angle that the parallel bundle makes with the axis in the object space. (This is the most general definition of focal length in geometrical optics, being the only one applicable to skew rays). The alternative method, in which the ray traced is initially parallel to the axis in the object space, is not as reliable, because the principal planes and the focal length are determined from the final slope of this ray in the image space.

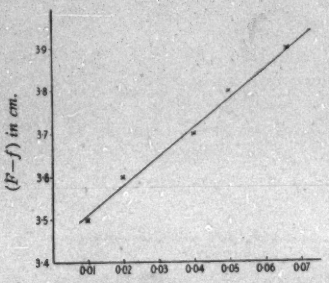


Figure 2.

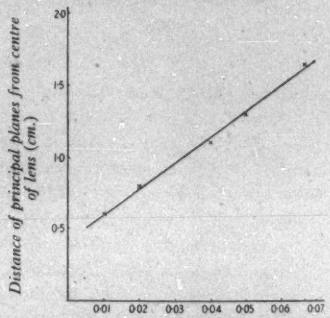


Figure 3.

For the lens described above, the distribution of the magnetic field along the axis was determined experimentally. The value of $f = 15$ cm. was chosen, and this example was fully traced with parallel bundles in the object and in the image spaces. These tracings are shown in figure 1. The actual focal length was found to be 18·9 cm. The principal planes were 1·65 cm. from the centre, and were " crossed ".

This fully-investigated case of $f = 15$ cm. showed that two parallel rays in the object space were sufficient to fix the cardinal points of the lens (since the cardinal planes are flat surfaces), and accordingly the lens was investigated for f equal to 20 cm., 25 cm., 50 cm. and 100 cm. The values of the corresponding focal lengths, and the positions of the principal planes, are given in table 1. Figure 2 shows how these may be plotted to give a straight-line graph. $(F-f)$ was plotted against $1/f$, and is seen to give a very good straight line. Also, figure 3 shows the distance of the principal planes from the centre of the lens

plotted against $1/f$. The linear relationship is again satisfactory. Two interesting features of these curves are that the differences $(F-f)$ between the true

Table 1

f (cm.)	$(F-f)$, the difference between the true focal length and f (cm.)	Distance of the focal point from centre of lens (cm.)	Separation of principal planes from centre of lens (cm.)
15	3·9	17·5	1·65
20	3·8	22·5	1·3
25	3·7	27·6	1·1
50	3·6	52·8	0·8
100	3·5	102·9	0·6

focal lengths F and the parameter f are nearly constant over a wide range of values of f; and that even for very large focal lengths there remains a definite displacement of the principal planes from the centre of the lens.

§ 4. THE SHORT MULTI-LAYERED SOLENOID

The calculations above apply only to a particular thick lens. The form of the results as given in figures 2 and 3 is clearly more general. By ray-tracing methods, the determination of the focal lengths and positions of the principal planes was effected for single-layered solenoids of lengths 20 cm. and various radii. From these figures, given in table 2, it is seen that these lenses have remarkably similar properties, and that the focal lengths described in terms of the f's remain constant over a large range of radii. Because of this property, the curves shown are sufficient to fix the optical properties of a very large variety of lenses wound in the form of solenoids. For if the length, radius and the parameter f for any coil are altered in the same ratio, the value of $(F-f)$ will be altered in this ratio, as will the separation of the principal planes from the centre of the coil; and these quantities can be found from table 2. This will be particularly easy as

Table 2. For a lens wound as a single-layered solenoid, 20 cm. long

Radius (cm.)	$(F-f)$ (cm.)	Distance of principal plane from centre (cm.)
10	4·0	1·5
5	3·9	1·4
2·5	2·4	0·3

the properties of the lens are not sensitive to changes in radii, and the results will still give a good approximation even if the ratio of the radii is not accurate.

Again, the same property allows the effects of the multiple layers with which

any practical coil would have to be wound to be taken into consideration. For unless the solenoid is wound with a very deep channel it will usually be sufficiently accurate to calculate f for the entire coil and to use the values of $(F-f)$, and the separation of the principal planes, for a mean layer of it.

§ 5. THE ACCURACY OF FOCUSING OF THE LENSES

The method of ray tracing described in § 3 was used above to determine focal lengths under conditions such that (*a*), (*b*), (*c*) and (*d*) were satisfied. Under these conditions it is clear that the " thickness " of the lens does not introduce any fresh aberrations (compared with a thin lens) of these types that can be described by tracing rays in the plane of the axis.

The formulae (10) and (11) used for ray tracing show that all rays initially parallel to the axis are brought to the same focus. Thus, under the specified conditions, the thick lens is free from spherical aberration. In fact, under these conditions, equation (3) becomes a linear homogeneous differential equation, and it follows that a point on the axis is accurately focused to a point on the axis. Again actual ray tracing of bundles both parallel to, and inclined at a definite angle to, the axis gives the same positions of the cardinal planes and the same values of the focal lengths.

According as the conditions (*a*), (*b*), (*c*) and (*d*) are not satisfied, inexactness is introduced into the focusing of the lens. It is not proposed to attempt to discuss here the aberrations produced by specific departures from these conditions, but in order to construct a lens of given quality it is necessary to see how to satisfy the conditions to the required degree of accuracy.

Conditions (*a*) and (*b*) must be satisfied for each point of a ray. Conditions (*c*) and (*d*) are integral conditions to be satisfied over the whole path of the ray. Thus, to determine the actual rotations of the different rays forming an image, (*c*) would have to be integrated along each ray. Alternatively, a differential method can be adopted to simplify the determination of the optical properties of a given magnetic field. Instead of considering, for example, the total rotation for a ray, the rate of rotation for an electron at each point of the field can be considered and the total rotation added up arithmetically. The above conditions for accurate image formation will now be considered in greater detail.

(a) *The constancy of v_x for all rays*

This condition fundamentally limits the available aperture for a permitted inaccuracy in the performance of any lens. Consider the formation of the image of a point on the axis which emits electrons of velocity v. The axial component of the velocity of an electron entering the lens inclined at an angle θ to the axis is $v_x = v \cos \theta$, compared with the value $v_x = v$ for an electron moving along the axis. Inside the lens, the ray will be bent round until at one point it has no radial component of velocity. Here, however, v_x will still be less than v, owing to the tangential velocity the electron has acquired. This

tangential velocity is given by equation (1). Suppose that at any point (x, r) the axial component of the magnetic field is H_x and the electron is moving parallel to the axis. The tangential velocity here is $\frac{r}{2}\left(\frac{e}{m}\right)H_x$, so that the axial velocity v_x is given by

$$v_x^2 = v^2 - \frac{r^2}{4}\left(\frac{e}{m}\right)^2 H_x^2. \qquad \ldots\ldots(13)$$

Let
$$\int_{-\infty}^{\infty} H_x^2 dx = H_x^2 l, \qquad \ldots\ldots(14)$$

where the length l defined is related to the thickness of the lens.

Then,
$$v_x^2 = v^2\left[1 - \frac{r^2}{4}\left(\frac{e}{m}\right)^2 \frac{\int H_x^2 dx}{lv^2}\right] \qquad \ldots\ldots(15)$$

or
$$v_x = v\left[1 - \frac{r^2}{2f}\cdot\frac{1}{l}\right]. \qquad \ldots\ldots(16)$$

Before the electron entered the lens it follows from simple geometrical considerations that

$$v > v_x \geqslant v\left[1 - \frac{r^2}{2f}\cdot\frac{1}{f}\right], \qquad \ldots\ldots(17)$$

according to the position of the point object, the sign of equality holding for the object placed at the object focal point.

These equations exhibit the striking superiority of a thick lens over a thin lens in this respect.* For the latter, $f \gg l$, and the angular aperture for a permitted variation in v_x is much less than that resulting from the cosine factor of the velocities of the rays entering the lens, but by winding the lens of such a length that $f \simeq 1$, apertures up to the full one fixed by the cosine factor can be used according to the position of the point object. Since, in the symmetrical lens, the principal points lie near the centre of the coil, l cannot conveniently be much greater than f. Moreover, due to the continuous bending of the rays in a thick lens, the value of r in equation (16) is smaller for a thick lens than for a thin one.

(b) *The form of the field*

The other conditions, (*b*), (*c*) and (*d*), depend on the form of the field, and can conveniently be considered together. The axially symmetrical magnetic field can be defined in the usual way by the potential

$$\Omega = A_0 - \frac{r^2}{2^2}\cdot\frac{\partial^2 A_0}{\partial x^2} + \frac{r^4}{2^2 4^2}\cdot\frac{\partial^4 A_0}{\partial x^4}, \qquad \ldots\ldots(18)$$

where A_0 is a function of x only.

* Indeed, it appears that the thinner the lens, the smaller is the aperture of the cone of rays accurately focused. A perfectly thin lens would not focus.

The condition $\frac{1}{2}\partial H_x/\partial x = -H_r/r$ shows that equation (1) would be strictly true if the axial component H_x of the field was everywhere given by

$$H_x = -A_0' + \frac{r^2}{8}A_0''' - \ldots . \qquad \ldots\ldots(19)$$

instead of by

$$H_x = -A_0' + \frac{r^2}{4}A_0''' - \ldots . \qquad \ldots\ldots(20)$$

Of course, except in certain trivial cases, the field given by equation (19) cannot exist, as it is inconsistent with Laplace's equation, but (19) is a good approximation to (20) provided that

$$\frac{r^2}{8}|A_0'''| \ll |A_0'|, \qquad \ldots\ldots(21)$$

and the accuracy with which (1) holds depends on the degree of accuracy with which this condition is satisfied.

Conditions (*c*) and (*d*) are actually integral conditions. They can, however, be replaced by the much simpler condition that H_x should at any point be independent of the distance of this point from the axis. In this form, the condition is too stringent, but any field which satisfies the simplified condition will certainly satisfy (*c*) and (*d*). For the axially symmetric field the new condition becomes

$$\frac{r^2}{4}|A_0'''| \ll |A_0'|. \qquad \ldots\ldots(22)$$

As the condition derived from (*b*) is less stringent than this, and as departure from exact fulfilment of both conditions produces errors in the same direction, the latter can be used alone.

For the examination of actual magnetic fields, the differential method of estimating the errors produced from these conditions is far simpler to apply. As far as the rotation of the image is concerned, the percentage variation in the rate of this, due to the electron being at the point (x, r) instead of being very close to the axis, is

$$100\frac{H(x,r)-H(x,0)}{H(x,0)}$$

$$= 25r^2\left|\frac{A_0'''}{A_0'}\right|. \qquad \ldots\ldots(23)$$

This method is worked out in figures 4 to 7. If

$$25r^2|A_0'''/A_0'| \leqslant n,$$

a region is defined in the lens by the inequality:

$$r \leqslant 1/r\sqrt{n}|A_0'''/A_0'|, \qquad \ldots\ldots(24)$$

where the error in the rate of rotation is everywhere $\leqslant n\%$. Such contours

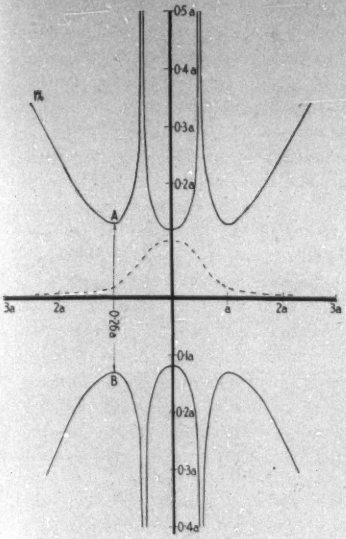


Figure 4. Circle of wire of radius a.

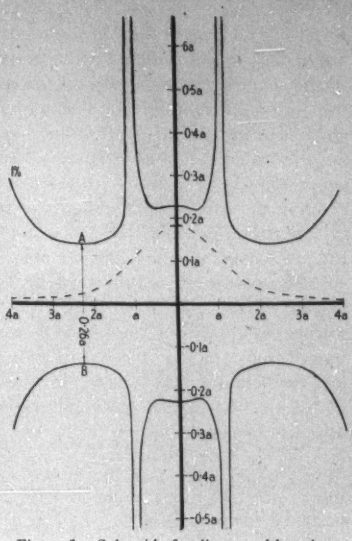


Figure 5. Solenoid of radius a and length a.

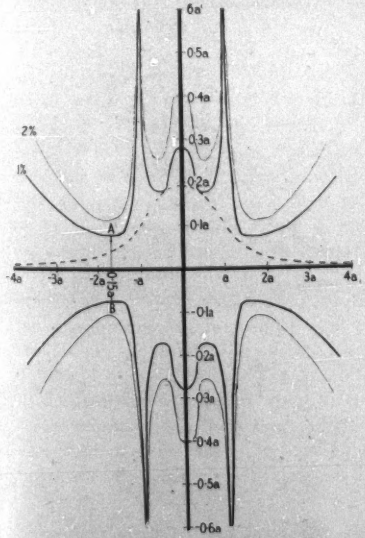


Figure 6. Solenoid of radius $a/2$ and length a.

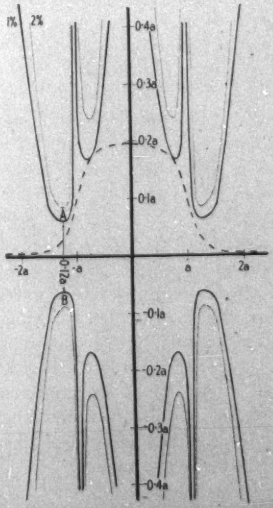


Figure 7. Solenoid of radius $a/4$ and length a.

The figures show the regions within which the electron tracks must be confined so that the deviations in their rates of rotation shall always be less than 1 %. The regions between the asymptotes produce deviations in the opposite directions to the regions outside them.

for errors of 1% are drawn for various fields. If the aberrationless ray tracks are drawn on top of these for the field which is to be used, the total variations in the rotation can be added up for each ray. The method is also useful for comparing different fields. The accompanying figures show that, although by winding a solenoid of small radius, the actual size of the entrance to the lens for a given permitted error is reduced compared with a solenoid wound with a larger radius, nevertheless this reduction is quite small. Further, any ray entering the lens passes through two regions where the errors in rotation are in opposite directions. The smaller the radius of the winding, the narrower and the closer these regions become. Further, they are pushed out into a position near the end of the coil where the actual field is weak, and an error in the rate of rotation contributes but little to the total rotation. Apart from the obvious advantages of cheapness, ease of winding, and gain in magnetic field derived from winding coils of small radii, the smaller the radius, the more rapidly does the field diminish outside the actual length of the winding of the solenoid. The merits of a particular winding can always be judged by drawing the aberrationless ray tracks on top of the error contours constructed for the winding as in figures 4 to 7, but it would appear that, except for special purposes, little advantage, if any, is to be derived from the use of solenoids of large radii.

§6. THE DESIGN OF A β-RAY SPECTROMETER

Attempts made hitherto by Klemperer (1938), Cosslett (1940) and others to use a magnetic lens as a β-ray spectrometer have all involved the use of thin lenses and the focusing of a point on the axis to form a point image by the method of focal isolation that was used years ago by R. W. Wood for his investigation of infra-red spectra. Recently Witcher (1941) has used a homogeneous magnetic field. To prevent unresolved rays along the axis from entering the receiver, they used a stop on the axis.

The obvious advantage to be derived from the use of thick lenses over thin lenses is that much weaker fields can be used. As Witcher points out, this becomes very important for the focusing of high-energy electrons. Thus to focus β rays of 15 Mev., Cosslett constructed a powerful electromagnet to give a focal length of about 40 cm. with a pole gap—i.e. lens thickness—of 5 cm. A solenoid wound for 50 cm. and used to give a focal length of 50 cm. would need a field along the axis of only 2000 gauss to focus these electrons. What is more important is that, according to the discussions of aberrations given here, such a lens should have far better properties.

The special feature of magnetic focusing is the rotation of the image with respect to the object. While for most of the applications of magnetic lenses this has only a nuisance value, it has been turned to specific use in a single-lens β-ray spectrometer constructed here in an attempt to increase the resolving power of these instruments.

Fundamentally, the method consists in forming a line image of a line source; the electrons are collected by a line receiver. The lens is designed to give a rotation of about 90°, and to be as free as possible from aberrations.

Very approximately for a solenoid, if

$$\left.\begin{aligned}\int H_x\,dx &= Hl,\\ \int H_x^2\,dx &= H^2 l,\end{aligned}\right\} \qquad \text{.....(25)}$$

then

equations which are sufficiently true for purposes of design. Thus to get the right rotation

$$\frac{1}{2}\left(\frac{e}{m}\right) Hl/v = \pi/2,$$

which fixes the value of the product Hl. Now the approximate focal length is given by

$$\frac{1}{f} = \frac{1}{4}\left(\frac{e}{m}\right)^2 \frac{H^2 l}{v},$$

and the value of H is chosen here so that the value of f obtained is somewhat less than l; for, as shown above, the lens is most free from aberrations when the length of the winding is as great as possible compared with the focal length. The lens actually constructed here has a length of winding of about 20 cm., to be operated at an f of 15 cm.

ACKNOWLEDGEMENTS

I should like to thank Professor P. M. S. Blackett, F.R.S., for the encouragement that he gave at the beginning of this work, and Professor J. D. Bernal, F.R.S. for his continued support. I owe much to my former colleague Dr. W. Ehrenberg, now at Electric and Musical Industries, Ltd., for many useful and encouraging discussions.

REFERENCES

Bush, H., 1926. *Ann. Phys., Lpz.*, **81**, 976; 1927. *Arch. Elektrotech.* **18**, 583.
Cosslett, V. E., 1940. *Proc. Phys. Soc.* **52**, 511.
Klemperer, O., 1935. *Phil. Mag.* **20**, 545.
Witcher, C. M., 1941. *Phys. Rev.* **60**, 32.

THE DETERMINATION, BY SMOKE MANOMETER, OF THE RATIO OF DENSITY OF TOWN GAS TO THAT OF AIR AT THE SAME TEMPERATURE AND PRESSURE *

By G. F. C. SEARLE, Sc.D., F.R.S.,
Cavendish Laboratory, Cambridge

Received 6 *March* 1942

ABSTRACT. The buoyancy effect of a column of coal gas is balanced against that of a column of hot air. At their upper ends, the columns are in free communication with the atmosphere. The lower ends of the tubes containing the gas and the hot air are joined to a smoke manometer; communication between the columns is established by turning a tap. The absence of drift of the smoke indicates that the pressures at the lower ends of the columns are equal. The effective height of the hot-air column is h cm., and the absolute temperature of the steam jacket surrounding it is θ'. The balancing height of the gas column is k cm.; the temperature of the gas is θ, the atmospheric temperature. Then, exactly,

$$\sigma/\rho = 1 - h(\theta' - \theta)/k\theta',$$

where the densities σ, ρ of gas and air are measured at a temperature and a pressure which are the same for both gas and air.

In a second method, the buoyancy effect of the gas is balanced against the defect of pressure due to the combination of a whirling arm and a static-head tube described in an earlier paper. If v be the velocity in its path of the test hole of the static-head tube and k be the height above the centre of the manometer of the upper (open) end of the gas column,

$$\sigma/\rho = 1 - v^2/2gk.$$

The value of σ/ρ may also be found by weighing an aluminium hot-water bottle (1) full of air and (2) full of gas, the temperature and pressure being the same in each case.

§ 1. THE SMOKE MANOMETER

The manometer is a modification of that shown in figures 6, 7 in a paper —quoted later as "Paper A"—by G. F. C. Searle (1941). Its details are shown in figures 1, 2, 3 below. A rectangular block of brass $3\frac{1}{4}$ inches in length and 1×1 inch in section is used. An axial duct $\frac{1}{8}$ inch in diameter is

* The experiments form good exercises for students and are described with that end in view. Unsteady movements of the air near the apparatus and variations of its pressure and temperature bar any claims of the methods to high precision.

bored from end to end. The taps R, S (figure 1) screwed into the ends of the block are coaxial with the duct. Coaxial with two other taps U, V are two holes meeting the duct at right angles. The plane RSUV is horizontal. The duct runs into H, a central vertical pit $^5/_{16}$ inch in diameter. From the centre of the face opposite to UV a horizontal hole J, $\frac{1}{8}$ inch in diameter, is bored into H. The outer end of J may be conical, as in figure 2, to allow free passage to the light. Figure 2 is a vertical section through the axis of J. A window W_1 of plane glass of good quality is set in a recess at the bottom of a larger recess turned out of the upper part of the block. A second window W_2 is set in a recess coaxial with J. The windows are secured with wax. Light from a 6-volt motor headlight lamp M is concentrated by a lens L and, when H contains smoke, a small volume of the smoke is brilliantly illuminated. The distance HL is 1·8 and

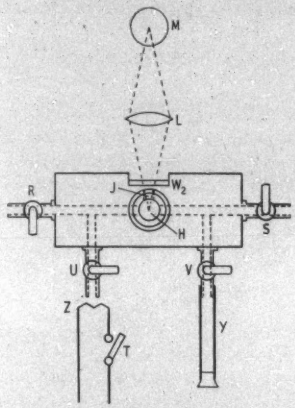


Figure 1.

Figure 2.

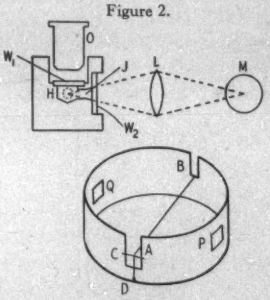


Figure 3.

Figures 1, 2, 3. Smoke manometer.

HM is 4·3 inch. The lamp and lens are mounted on the base which carries the manometer block. A vertical microscope, with objective O, is used for viewing the smoke. The Brownian movement is well seen. The microscope is fixed to a carriage which slides on two vertical rods springing from the base. Its magnifying power is 40; a greater power—say 80—would be better.

When smoke is to be drawn into the manometer, taps R, S are closed and U, V are opened. A small cautery Z, which is heated by a current when the switch T is closed, is placed near U, and a match-stalk is held in contact with Z. A rubber tube Y, closed at one end, is fixed to V. If Y be squeezed and then released, smoke is drawn into the manometer duct. Then U, V are closed.

If the pressure in a tube leading to R be greater than in a tube leading to S, the smoke will drift in the direction from R to S, when R and S are opened.

A difference of pressure of 0·1 dyne cm.$^{-2}$ is easily detected. A difference of 2 or 3 dyne cm.$^{-2}$ quickly sweeps the smoke out of the duct, if the taps be fully open.

Messrs. Rotherham and Sons, Ltd., Coventry, supplied their No. 54 taps. These have a bore of $\frac{1}{8}$ inch and have been tested to 75 lbs. inch^{-2} of air pressure. It is essential that the taps and the windows should not leak.

A "cross-wire", by which the drift of the particles may be judged, is stretched across a short length of tube (figure 3) fitting into H. A slot A, corresponding to J, allows passage to the light. Diametrically opposite is a slot B of suitable length. A fine wire C is soldered across A at the height of the axis of J. One end of a silk fibre is secured by wax at D; the fibre passes over C and goes to the bottom of slot B, and the other end is secured with wax as at D. Flats filed on the outer surface of the tube at A and B leave space for the wire C and for the wax. When illuminated, the fibre appears as a bright line. A pair of openings P, Q, diametrically opposite each other, correspond to the duct from R to S.

§ 2. THEORY OF BALANCING COLUMNS

Let ABC, DE be two narrow tubes, bent as in figure 4. Let C, E be at the same level as the horizontal duct of the smoke manometer M, with which the tubes communicate through the taps R, S. The ends A, D are open to the atmosphere. The medium in DE is coal gas of the town supply, at atmospheric

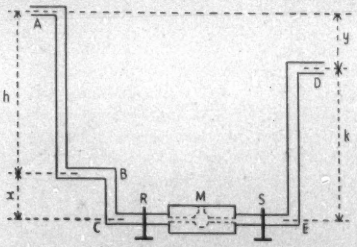


Figure 4. Diagram of balancing columns.

absolute temperature θ, that in the non-horizontal part of AB is air heated to θ' by a steam jacket surrounding the effective part of AB, and that in BC is air at temperature θ. The conditions in the horizontal parts of the tubes are governed by the principle that, if the medium be at rest, a *horizontal* line, curved or straight, drawn in the medium and passing, in different parts of its length, through hot or cold air or through gas or a mixture of air and gas, is a line of equal pressure. The air tube is shown with an extra piece BC to meet the case where it is not convenient to have B at the same level as R.

Let [A]$-$[B], the height of A above B, be h, let [B]$-$[C]$=x$, [A]$-$[D]$=y$ and [D]$-$[E]$=k$.

If ρ and σ be the densities of air and of gas at pressure p and absolute temperature θ, we have $\rho = p/R\theta$, $\sigma = p/S\theta$, where R, S are constants. If z be measured downwards, we have, in air at temperature θ,

$$dp = g\rho dz = gpdz/R\theta.$$

Hence $dp/p = gdz/R\theta$, and thus, if θ be independent of z and if p_0 correspond to $z = 0$,

$$\log p - \log p_0 = gz/R\theta. \qquad \ldots\ldots(1)$$

When we apply (1) to the air in AB and BC, we find

$$\log p_B - \log p_A = gh/R\theta', \quad \log p_C - \log p_B = gx/R\theta,$$

and thus

$$\log p_C - \log p_A = g(h/R\theta' + x/R\theta). \qquad \ldots\ldots(2)$$

For the air between A and D and for the gas in DE we have

$$\log p_D - \log p_A = gy/R\theta, \quad \log p_E - \log p_D = gk/S\theta,$$

and thus

$$\log p_E - \log p_A = g(y/R\theta + k/S\theta). \qquad \ldots\ldots(3)$$

When the manometer is balanced, $p_E = p_C$; then, by (2) and (3),

$$k/S\theta + y/R\theta = h/R\theta' + x/R\theta. \qquad \ldots\ldots(4)$$

Now $x + h = y + k$, and thus $x - y = k - h$. Then

$$\frac{R}{S} = \frac{\theta}{k}\left\{\frac{x-y}{\theta} + \frac{h}{\theta'}\right\} = \frac{\theta}{k}\left\{\frac{k}{\theta} - \frac{h}{\theta} + \frac{h}{\theta'}\right\}$$

$$= 1 - \frac{h}{k}\cdot\frac{\theta' - \theta}{\theta'}. \qquad \ldots\ldots(5)$$

This exact result does not involve logarithms.

The constants R, S are independent of temperature and pressure. If we specify that ρ and σ are measured at a temperature and a pressure which are the same for both air and gas, we have $\sigma/\rho = R/S$, and then, by (5),

$$\frac{\sigma}{\rho} = 1 - \frac{h}{k}\cdot\frac{\theta' - \theta}{\theta'}. \qquad \ldots\ldots(6)$$

Figure 4 may suggest that the main parts of the tubes are vertical. But they need not be, and are not, vertical; it is only differences of level that count.

§ 3. THE APPARATUS

The general arrangement is shown diagrammatically in figure 5. Free communication at D between coal gas and atmosphere is essential, but, unless it be suitably controlled, the column of gas will be contaminated with air. The outlet from gas to atmosphere is a small hole, centre D, in the end of the tap F. The tap is fixed to one end of the chamber G, a horizontal tube 24·5 cm. in length and 1·7 cm. in diameter. At the other end is a nipple for a rubber tube T_1. By the principle stated in § 2, a horizontal line passing from the atmosphere

through D into G is a line of constant pressure. The large volume of G ensures that contamination by air of the gas in T_1 is very slow.

The tube T_1 joins G to the arm S_1 of a three-way piece. The arm S_2 is joined by a short rubber tube to the corresponding nipple of the manometer M. The centre of the hole in this nipple is E. The third arm is connected with Q, the gas supply tube, through the tap S_3.

If the taps S_3, F be open, gas will fill T_1 and G and will stream out at D. In 10 or 15 seconds, all the air will be expelled, and then F is closed. When the manometer test is to be made, tap S_3 is closed and F is opened.

The greatest height of D above E is about 50 cm. If D were vertically above E, the chamber G would be in the way of the observer when he uses the microscope. A plane surface, e.g. a strip of plate glass, provided with screws, is levelled and the heights of D and E above this surface are measured. If these be d and e, then $k = d - e$.

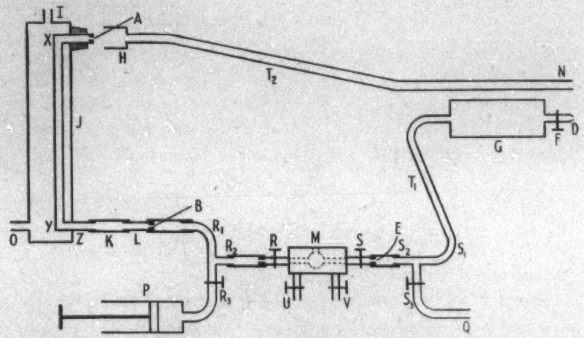


Figure 5. Buoyancy-tube method.

The chamber is supported by a carriage which slides on and can be clamped to a vertical rod. This provides a coarse adjustment of height. A fine adjustment relative to the carriage is effected by a screw.

The greater part of the bent tube AXYZ is surrounded by a steam jacket J, which has efficient lagging.* Steam enters at I and leaves by the tube O, which is coaxial with the part YZ of the inner tube. At the end of XA is a small hole of centre A. The tube YZ is prolonged by a rubber tube K and a metal tube L; both tubes are horizontal. At the end of L is a small hole of centre B. The rubber tube K ensures that L is very nearly at atmospheric temperature. Tube L is connected by a rubber tube to the corresponding nipple of the manometer M through the arms R_1, R_2 of a three-way piece. The third arm, in which is a tap R_3, leads to a small pump P. By P, air can be driven through BYXA to

* A full description is given in paper A cited in § 1.

clear out any coal gas which, through unskilful manipulation, may have penetrated M and entered the tube R_2A. The tap R_3 is closed when the balance is being found.

Tubes O, Z act as trunnions and turn in bearings attached to a length of ⊥-girder. The height of A above B, viz. h, is varied by turning the tube about OZ as axis. The edge of the girder's fin, which has been planed, is levelled. An adjustable strut rising from the girder supports the tube at any desired slope. The heights of A and B above the levelled edge are a and b, and $h=a-b$.

A horizontal line drawn through B and lying within LY is a line of equal pressure, and similarly for AX. Thus, if XA, YB be horizontal, $p_X=p_A$ and $p_Y=p_B$, and the theory of § 2 is applicable.

The temperature of the air near the apparatus is found by a thermometer for each comparison of k with h. The temperature may rise 2 or 3 degrees in half an hour in a class-room. The thermometer should be so placed that it is not heated by radiation from the boiler or by radiation from or by contact with any object, such as the table, which is heated by the boiler. The air itself is little heated by radiation, and slight draughts tend to minimize local rises of air temperature.

§ 4. ANTI-DRAUGHT DEVICE

The apparatus is sensitive to draughts. If, for instance, taps R, S be closed and U, V be open, a to-and-fro movement of a finger near the end of U, or even near the end of a tube a metre in length attached to U, will cause a corresponding movement of the smoke. In a busy class, the movements of students cause draughts which affect the tube A and the tap F unequally, and so lead to spasmodic drifts of the smoke. The difficulty is overcome by a simple device. A short piece of rubber tube fits on the nipple A of the heated tube. One end of a horizontal tube H, of metal or, better, of fibre or other poor conductor, can fit over the rubber; to the other end of H is joined a rubber tube T_2 which leads to an open nozzle N near the tap F of G. The axis of N is parallel to that of F and the ends of N and F are not more than 2 or 3 cm. apart. The part of T_2 near H is horizontal for a little distance, and care is taken that the remainder of T_2 is at atmospheric temperature. Then, when H is in position on A, the pressure at A inside H equals that of the atmosphere at the same level, unless there be large streaming motions of the air in the room. The device is so effective that violent local aerial disturbances near N and F produce only minute and transient movements of the smoke.

§ 5. TEMPERATURE OF STEAM

The temperature T° c. of the steam in the jacket can be found from a steam table. This gives T as a function of the pressure as measured by a barometer, with its mercury at 0° c., at sea level in latitude 45°, where $g_N=980{\cdot}62$ cm. sec.$^{-2}$ At the Cavendish Laboratory, $g_L=981{\cdot}27=1{\cdot}000663g_N$. If μ_1 and μ_0 be the density of mercury at t_1° c. and at 0° c., $\mu_1=\mu_0(1+0{\cdot}000182t_1)^{-1}$. If a pressure,

which gives height H_N cm. in latitude 45° with mercury at 0° c., give H cm. at the Laboratory with mercury at t_1°c. we have $g_N\mu_0 H_N = g_L\mu_1 H$, and thus

$$H_N = H(1+0{\cdot}000663)(1+0{\cdot}000182t_1)^{-1} = H + H(0{\cdot}000663 - 0{\cdot}000182t_1).$$

In the small term, we put $H = 76$. Then

$$H_N = H + 0{\cdot}050 - 0{\cdot}014t_1.$$

For H_N near 76 cm., the boiling point rises by 0·37° c. when H_N rises by 1 cm., and $T = 100 + (H_N - 76) \times 0{\cdot}37$. Thus

$$T = 100 + (H - 75{\cdot}95) \times 0{\cdot}37 - 0{\cdot}0052t_1. \qquad \ldots\ldots(7)$$

If the temperature of the air near the apparatus be t° c., the absolute temperatures of the air and of the steam are $\theta = t + 273{\cdot}2$, $\theta' = T + 273{\cdot}2$.

§ 6. EXPERIMENTAL DETAILS

The boiler, containing water, is started. The girder and the glass plate are levelled and b and e are measured. The height a of the upper nipple of the hot-air tube above the girder is found, and then the anti-draught device is adjusted. When steam is passing freely, the cell is charged with smoke and taps U, V are closed. Gas is allowed to flow through S_3 and F, and thus the chamber G is filled with gas. Then S_3 and F are closed. When a test for balance is to be made, F is opened. Then, R being open, S is opened cautiously; the direction of drift shows whether G should be raised or lowered. When the balance has been found, the height d of D above the plate is measured. If, when one pair of values of a–b or h and d–e or k has been found, k be set against h on a slide rule, approximate values of k can be read off for other values of h, and thus time can be saved in reaching the desired balances. The values of h should be large so as to minimize the effects of zero errors.

For each balance of k against h, the air temperature is observed. The boiling point T° must also be known, but, as T rises by no more than 0·1° for a rise (0·27 cm.) of H_N from 76 to 76·27 cm. or for a fall of 19° in t_1, the temperature of the mercury, the variations of T during the experiment are usually negligible.

A test was made of the efficiency of the chamber G (figure 5) in preventing contamination of the gas column by air. The effective height h of the hot-air column was 83·45 cm. The tube T_1 and the chamber G were filled with gas, and then the values of k needed for a balance were found at various times after the filling. The tap F was kept open throughout. The changes with time of the apparent value of σ/ρ are shown below.

Time (min.)	0	1	3	5	7	9	13	17
$10\,\sigma/\rho$	5·02	5·02	5·04	5·06	5·10	5·11	5·21	5·32

The results show that no error due to contamination with air need be feared if the balance be found within two minutes of filling the chamber with gas. The test was made a few days before the measurements recorded in § 7. The ratio σ/ρ varies from day to day. At the Gas Works, σ/ρ is found by the method of effusion through a very small orifice.

§ 7. PRACTICAL EXAMPLE

Miss Mary Stead, of Girton College, made the observations. Conditions were favourable. The temperature of the air near the apparatus was steady, and no disturbance due to draughts was observed; the anti-draught device was not used. The barometric height was initially 76·91 cm.; the effect on the boiling point of a slight fall in H was negligible. The temperature of the barometer was 13·9° c. By (7),

$$\theta' = 273{\cdot}2 + 100 + (H - 75{\cdot}95) \times 0{\cdot}37 - 0{\cdot}0052 \times 13{\cdot}9 = 373{\cdot}5° \text{ K}.$$

For each h, four independent settings and readings for k were made. The mean divergence, without regard to sign, from the mean k, of the four values of k found for a given h is denoted by Δ. The values of $1000\Delta/k$ in the table are small.

h	k (mean)	$\frac{1000\Delta}{k}$	θ	$\frac{\theta' - \theta}{\theta'}$	$\frac{h}{k}$	$\frac{\sigma}{\rho}$
(cm.)	(cm.)		(° K.)			
89·70	41·300	0·61	287·2	0·23106	2·1719	0·4982
87·00	39·825	0·63	287·3	0·23079	2·1846	0·4958
81·15	37·262	1·01	287·3	0·23079	2·1778	0·4974
77·05	35·338	1·24	287·3	0·23079	2·1804	0·4968
72·60	33·412	1·12	287·3	0·23079	2·1729	0·4985

The mean value of σ/ρ is 0·49734, with a " probable error " of 0·00033.

§ 8. WHIRLING-ARM METHOD

In place of the column of hot air, the whirling arm described in §§ 3, 4 of Paper A may be used to give a defect of pressure. In figure 6, BC is the arm, which turns about the vertical axis OZ.

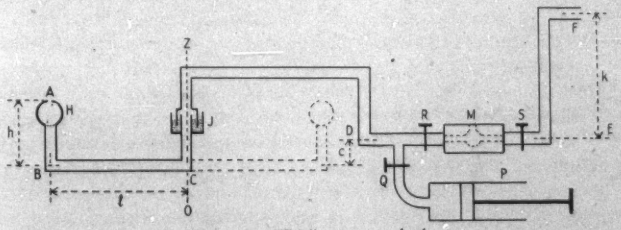


Figure 6. Whirling-arm method.

The distance of the test hole A in the static-head tube H from OZ is l, and its distance above BC is h. There is a liquid joint J in CD. The line DE through the smoke manometer M is horizontal, and BC is at distance c below DE. The tube EF contains coal gas; the distance of F above DE is k. The details of the arrangement on the gas side are shown in figure 5. The pump P of figure 5 is connected, through the tap Q, to the tube from D to M, and serves to expel any trace of coal gas from AD.

If the absolute temperature of the apparatus and of the surrounding air be θ, we have, by § 2,

$$\log p_B - \log p_A = gh/R\theta, \qquad \log p_C - \log p_D = gc/R\theta. \quad \ldots\ldots(8), (9)$$

Let ω be the angular velocity of BC about OZ. Then in BC

$$dp/p = \rho r\omega^2 dr/p = r\omega^2 dr/R\theta,$$

and thus

$$\log p_B - \log p_C = \tfrac{1}{2}l^2\omega^2/R\theta = \tfrac{1}{2}v^2/R\theta, \quad \ldots\ldots(10)$$

where v is the velocity of the test hole A along its path. From (8), (9), (10),

$$\log p_D - \log p_A = [g(h-c) - \tfrac{1}{2}v^2]/R\theta. \quad \ldots\ldots(11)$$

If the static-head tube H be suitably corrected (Paper A, § 4), the pressure at A inside H equals that in the atmosphere at a distance from A and at the same level as A.

On the other side of the manometer, we have, for the air and the gas respectively,

$$\log p_A - \log p_F = g(k+c-h)/R\theta, \quad \log p_E - \log p_F = gk/S\theta, \quad \ldots\ldots(12), (13)$$

and thus

$$\log p_E - \log p_A = g(h-c-k)/R\theta + gk/S\theta. \quad \ldots\ldots(14)$$

When the manometer shows a balance, $p_E = p_D$. When we subtract (14) from (11), and then multiply by θ, we have

$$gk/R - gk/S - \tfrac{1}{2}v^2/R = 0,$$

and thus

$$\frac{R}{S} = 1 - \frac{v^2}{2gk}. \quad \ldots\ldots(15)$$

This equation takes the place of (5) in § 2. If ρ, σ be measured at a temperature and a pressure which are the same for both air and gas, $\sigma/\rho = R/S$, and thus

$$\frac{\sigma}{\rho} = 1 - \frac{v^2}{2gk}. \quad \ldots\ldots(16)$$

No knowledge of the temperature is needed in (16), but uniformity of temperature is assumed.

The use of the whirling arm involves the correction of the static-head tube required for the small velocities employed, and this correction has to be made by running the whirling arm against a heated buoyancy tube, as described in § 11 of Paper A. The anti-draught device (§ 4) cannot be employed with the whirling arm. For these reasons, it seemed best in the present paper to treat the method of the buoyancy tube as standard rather than that of the whirling arm.

§ 9. DETERMINATION OF σ/ρ BY WEIGHING

The results found by the previous methods may be tested by weighing. A vessel of mass M and internal volume V is weighed when it is filled with (1) air and (2) coal gas at the temperature and pressure of the surrounding atmosphere. In (1) the mass of the vessel and its contents is $M+\rho V$, and in (2) is $M+\sigma V$. The balancing load may consist of brass weights and aluminium weights. We

suppose that the brass weights are the same in each case, and that the balancing load for (1) exceeds that for (2) by D gm. of aluminium of density α. The effective mass of D gm. of aluminium is $D(1-\rho/\alpha)$, where ρ is the density of the air in the balance case. The buoyancy effects of the air on the vessel and on the brass weights are the same in both cases. Hence

$$(\rho-\sigma)V=D(1-\rho/\alpha),$$

and thus

$$\frac{\sigma}{\rho}=1-\frac{D}{V}\left(\frac{1}{\rho}-\frac{1}{\alpha}\right). \qquad \ldots\ldots(17)$$

An aluminium hot-water bottle B (figure 7), of about 1500 c.c., may be used. To its mouth is fixed, by marine glue or other adhesive, a plate C through which pass, with soldered joints, two tubes L and S. Tube L goes nearly to the bottom of B, and S only just passes through C. The tubes project a little distance above C. Corks are provided for them. The temperature of B is assumed to be that of the atmosphere, and hence it should not be touched by the hand. It

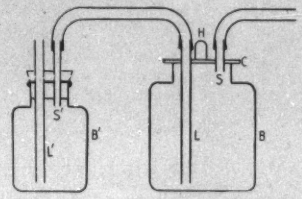


Figure 7. Weighing method.

may be lifted by a rod inserted into a loop H soldered to C. Care must be taken that the bottle does not pick up dirt of any kind.

For the air filling, a tube from the blowpipe bellows (not from the observer's mouth—Oh, dear, what things they do!) is fixed to L, and the bellows are worked until it is certain that any gas remaining after a previous gas filling has been expelled. Then the bottle, with the corks, is put into the balance case; the tubes L, S are left open. After a few minutes to allow the bottle and its contents to attain the temperature of the air in the case, the weight is found—best by the method of oscillations.

The bottle is now filled with gas. A tube leading from the gas supply is attached to S, and gas is allowed to flow for a little time to expel all air from B. If the gas come from a main near heating pipes, its temperature may be above that of the room. If L and S be corked immediately after the tube is removed from S, the pressure of the contained gas will be below that of the atmosphere when its temperature has sunk to that of the atmosphere, or, if the corks leak, air will flow in as the gas cools. To meet this difficulty, a bottle B′, with tubes L′ and S′, is put in tandem with B, as in figure 7; tube S′ is connected to L. After gas has streamed through B and B′ for a sufficient time, the tube from the

supply pipe is removed from S, which is immediately corked. The pair of bottles is then moved to a place near the balance case. If the gas cool, air will flow into B′ through L′; but the small volume of gas which flows into B from B′ will be only slightly contaminated with air. After a few minutes, the tube connecting S′ to L is removed, and L is immediately corked. Then the bottle B, with the contained gas, is weighed.

Before the plate C is fixed to B, the volume V_0 of B to the level of the top of the neck is found by weighing it (1) empty and (2) filled with water at a known temperature. Allowances are made for the buoyancy of the enclosed water and of the balancing masses. The coefficient of cubical expansion of aluminium ($7{\cdot}65 \times 10^{-5}$) is so small that changes of volume due to moderate changes of temperature are negligible. The brass tubes L and S are weighed before they are soldered into C, and the volume V_1 of brass below C is estimated. The small volume of the tubes between the corks and the level of the lower face of C is V_2. Then $V = V_0 - V_1 + V_2$.

REFERENCE

SEARLE, G. F. C., 1941. "A whirling arm for producing small differences of pressure, its calibration by aid of a smoke manometer and its application to measurements of surface tension." *Proc. Phys. Soc.* **53**, 681.

DEMONSTRATION

PHOTOMICROGRAPHS illustrating the paper "An achromatic reflection microscope for use with visible or ultra-violet light" (*Proc. Phys. Soc.* **53**, 714 (1941)). *Demonstrated* 6 *February* 1942 *by* B. K. JOHNSON.

IN the above paper, the writer described an achromatic reflection microscope and gave details of the computed aberrations of the optical system. Photographs taken with the completed instrument were submitted with the original paper, but were not published at the time; an opportunity was, however, given to show them at a later meeting, and (because of the interest which they aroused among those Fellows able to attend) the author was subsequently invited to prepare a short description of the demonstration, illustrated by the photomicrographs in question. This description follows, though the original text should be consulted in order to appreciate the full meaning of each of the photomicrographs.

Figures 1 and 2 illustrate the definition obtainable with the reflection microscope designed for use with monochromatic light (λ 5461 A.) employing glass for the optical system (see *J. Sci. Instrum.* **11**, 384 (1934)). The instrument was originally intended for the observation of opaque objects in order to overcome back-reflection defects, but it may be used for transparent objects as shown in these two photographs.

Figures 3 (*a*), (*b*) and (*c*) show an opaque object (CuS) photographed with the achromatized reflection microscope, but using light of wave-lengths 5461 A.,

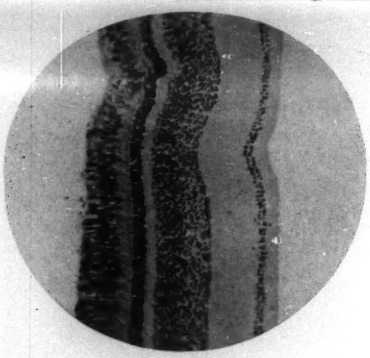


Fig. 1. Retina of owl.
λ=5461A. N.A.=0·31.
M=105x.

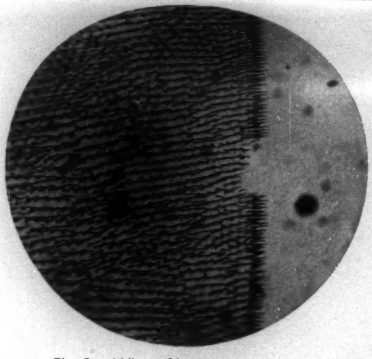
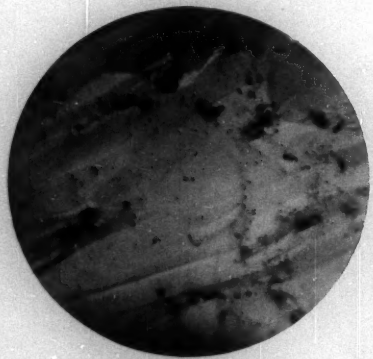


Fig. 2. Wing of insect.
λ=5461A. N.A.=0·31.
M=105x.

(a) (b) (c)



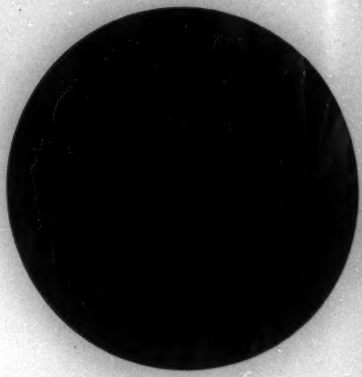
λ=5461A.

λ=3651A.

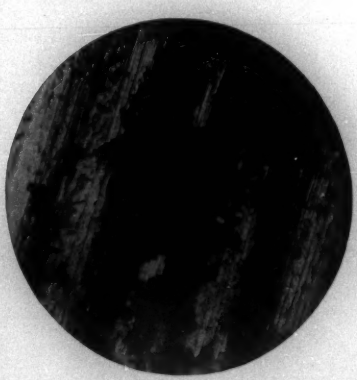
λ=2749A.

Fig. 3. Differential reflection effects when using various U.V. wavelengths. Object: CuS. N.A.=0·47. M=140x.

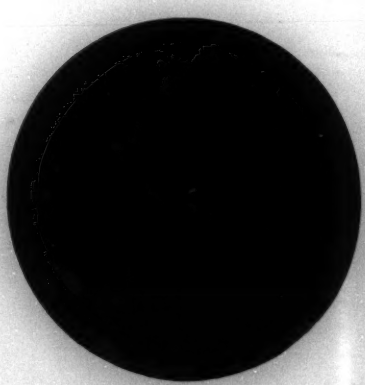
Specimen (zinc blend, Tsumed, S.W. Africa), illuminated and photographed with white light.
M=140x. N.A.=0·47.
Fig. 4.



Same specimen, illuminated by light of 3650 to 3131A and photographed in colour by the fluorescent light emitted.
M=140x. N.A.=0·47.



Specimen (an aluminium alloy) illuminated and photographed with white light.
M=140x. N.A.=0·47.
Fig. 5.



Same specimen, illuminated by light of 3650A to 3131A and photographed in colour by the fluorescent light emitted.
M=140x. N.A.=0·47.
N.B. A filter (transmitting no light below 4000A) was mounted in the plane of the exit-pupil of the eyepiece. Exposure=75 hours

Figs. 4 and 5. Comparison photomicrographs taken with B.K.J. reflection microscope showing benefit of fluorescent effects made available by the optical system of the instrument.

3651 A. and 2749 A., respectively. In this illustration there are two points worthy of note: firstly, no alteration in the fine adjustment of the microscope has been made for the three photomicrographs, which it will be seen are in equally good focus. Secondly, the effects of differential reflection (when using light of various ultra-violet wave-lengths) can be clearly seen.

Figures 4 and 5 show the possibilities of improving contrast and of detecting impurities by utilizing fluorescent effects.

In each pair of photomicrographs the monochrome picture gives the effect seen when white light only is used, whilst the coloured illustration shows the appearance of precisely the same part of each object when it is illuminated by ultra-violet light of wave-length 3650 A. to 3131 A. and seen by the fluorescent light emitted by the specimen. These photomicrographs in colour are only intended to convey what the eye actually sees when looking into the microscope, and it is not suggested that such photographs should be taken for all specimens which fluoresce, for the exposure required in this case (namely 75 hours) would in general be prohibitive. They do serve to show, however, that here is an additional means of aiding the possible identification of the components of an alloy (for example), or in the detection of impurities; and that the instrument described in the original paper is well suited to this purpose.

DISCUSSION

On papers by W. E. BENHAM (*Proc. Phys. Soc.* **54**, 121 (1942)) and G. D. YARNOLD (*Proc. Phys. Soc.* **54**, 47 (1942)) entitled "On the nature of temperature" and "Notes on electric and magnetic dimensions", respectively.

Professor DINGLE. I think Mr. Benham is in error in assuming that dimensions are associated with the *nature* of temperature: they characterize the *method of measuring* temperature, and this is arbitrary. Thus, if we define temperature as energy per unit mass, temperature will have finite dimensions in length and time, because a change in the units of measurement of length and time will cause a change in the measured value of a temperature. On the other hand, if we define temperature as a thermometer reading, the degree being 1/100 of the relative expansion of a standard substance between the temperatures of melting ice and boiling water, then the measured value of a temperature is unaltered by a change in any other unit, so that temperature has an independent dimension of its own. There is no meaning in asking what are the "true" dimensions until we have decided which is the "true" method of measuring temperature.

Much confusion is caused by the tacit assumption that the two methods are equivalent. An outstanding instance of their divergence is found in the temperature of interstellar space. Expressed as energy per unit mass, the temperature is some thousands of degrees, but a thermometer placed there

would read about 3° absolute. The molecules move very fast, but the matter is so rarefied that the energy which the thermometer receives by impact is radiated as fast as it is received.

There is a similar confusion in the measurement of time. If the unit of time is the mean solar second, or sidereal day, the measurement of time is unaffected by the measurement of anything else, and time has dimensions of its own. The Earth then rotates uniformly by definition. Astronomers, however, now believe that the Earth is "slowing down". This means that they must tacitly adopt another standard of time, which on investigation proves to be the standard according to which Newton's first law of motion (with light substituted for an undisturbed body) is automatically true. On this standard equal times are those in which light moves over equal lengths, and a change in the measurement of length therefore affects the measurement of time; hence time has finite dimensions in length.

Much of the confusion in this subject arises from oversight of the fact that dimensions characterize the arbitrary method of measuring physical quantities, and not the "ultimate nature" of the quantities.

Dr. E. A. Guggenheim. The object of the following remarks is to indicate how a little extra care in our choice of terminology could bring to an end the hitherto ceaseless confusion between quantities with physical dimensions on the one hand and pure numbers on the other. A quantity with physical dimensions is one whose numerical measure depends on the units used, whereas a pure number is obviously independent of these units. This statement can be illustrated by two familiar examples.

If we define density as mass per unit volume, then the density of water is (in round figures) 1 kg./l. or 62½ lb./ft.^{3} Again, if we define heat capacity as the energy required (under certain specified conditions) to produce unit rise of temperature in unit quantity of a substance, then the heat capacity of water is (in round figures) 1 B.Th.U./lb. deg. F., or 4·2 joules/g. deg. C., or 18 cal./mole deg. C. If, further, we define specific gravity as the ratio of the density of a substance to that of water, then the specific gravity of mercury is 13·6. Again, if we define specific heat as the ratio of the heat capacity of a substance to that of water, then the specific heat of mercury is 0·033. According to the chosen definitions it is evident that density and heat capacity have physical dimensions, while specific gravity and specific heat are pure numbers.

Quantities with physical dimensions, as well as pure numbers, may conveniently be denoted by symbols. For example, if $\mathbf{c}$ denotes the speed of light in empty space, then (in round numbers) $\mathbf{c} = 3 \times 10^8$ m./sec. = 186,000 miles/sec. = 1 light-year/year, and for $\mathbf{c}$ so defined it is wrong to write $\mathbf{c} = 3 \times 10^8$ (or any other number).

Turning now to electromagnetic quantities, if f denotes the force of repulsion between two electric charges q_A and q_B in free space at a distance apart r_{AB}, then

according to Coulomb's law the ratio $q_A q_B / f r_{AB}^2$ is a universal constant, which we can denote by $\boldsymbol{a}$. A short time ago I suggested (Guggenheim, 1941)* that the charge which repels an equal charge at a distance of 1 cm. in empty space with a force of 1 dyne be called a *franklin*, and I shall use this terminology. We have, then, for the universal constant $\boldsymbol{a}$ (in round numbers),

$$\boldsymbol{a} = 1 \frac{\text{franklin}^2}{\text{erg cm.}} = \frac{1}{9\times10^9} \frac{\text{coulomb}^2}{\text{joule m.}} = \frac{1}{9\times10^9} \frac{\text{sec.}}{\text{ohm m.}},$$

and it is evident that $\boldsymbol{a}$ has physical dimensions.

If we denote current elements by $i\,ds$, magnetic poles by p and the velocity of light by $\boldsymbol{c}$, all these quantities being measured in any units whatever, then we can construct table 1 for forces of attraction in free space and in a medium. The

Table 1. Forces of attraction between electric charges, current elements and magnetic poles

	Electric charges	Current elements	Magnetic poles
Empty space	$\frac{q_A q_B}{\boldsymbol{a} r_{AB}^2}$	$\frac{i_A i_B ds_A ds_B}{\boldsymbol{a}\boldsymbol{c}^2 r_{AB}^2}$	$-\frac{\boldsymbol{a}\boldsymbol{c}^2 p_A p_B}{r_{AB}^2}$
Medium	$-\frac{q_A q_B}{\epsilon \boldsymbol{a} r_{AB}^2}$	$\frac{\mu i_A i_B ds_A ds_B}{\boldsymbol{a}\boldsymbol{c}^2 r_{AB}^2}$	$-\frac{\boldsymbol{a}\boldsymbol{c}^2 p_A p_B}{\mu r_{AB}^2}$

quantities ϵ and μ are characteristic constants of the medium and, since by the chosen definition each is a measure of the ratio of a force in empty space to a force inside a medium, it is evident that each is a pure number.

In order to avoid ambiguity between the quantities $\epsilon\boldsymbol{a}$ and $\mu/\boldsymbol{a}\boldsymbol{c}^2$, both having physical dimensions, on the one hand, and the pure numbers ϵ and μ on the other, I suggest the adoption of the terminology shown in table 2. In choosing

Table 2. Proposed terminology

$\epsilon\boldsymbol{a}$ Permittivity of medium	$\frac{\mu}{\boldsymbol{a}\boldsymbol{c}^2}$ Permeability of medium
ϵ Dielectric constant of medium	μ Magnetic constant of medium
$\boldsymbol{a}$ Permittivity of empty space	$\frac{1}{\boldsymbol{a}\boldsymbol{c}^2}$ Permeability of empty space
$= 1 \frac{\text{franklin}^2}{\text{erg cm.}} = \frac{1}{9\times10^9} \frac{\text{sec.}}{\text{ohm m.}}$	$= \frac{1}{9\times10^{20}} \frac{\text{g. cm.}}{\text{franklin}^2} = \frac{1}{10^7} \frac{\text{ohm sec.}}{\text{m.}}$

* References are on p. 296.

this terminology I have been guided by a desire for the maximum precision consistent with the minimum departure from current practice. It will be noticed that the only essentially new terms are the *franklin* and the *magnetic constant*.

Attention has been drawn to the physical dimensionality of the universal constant $\boldsymbol{a}$ in a very instructive paper by Sommerfeld (1935), unfortunately too little known. Sommerfeld used the symbol ϵ_0 where I have used $\boldsymbol{a}$ and ϵ where I have used $\epsilon\boldsymbol{a}$. I think that the notation which I propose has two advantages over that of Sommerfeld's paper. Firstly, I propose that a single symbol $\boldsymbol{a}$ be used for a universal constant: secondly, I propose that the quantities usually listed in tables be also denoted by the single letters ϵ and μ. The greatest change entailed by my proposals is that tables of μ should be renamed tables of *magnetic constants*.

In conclusion, I want to emphasize that, whereas the question whether an unambiguously defined quantity x has physical dimensions or is a pure number has only one answer; when this answer is "Yes", the further question "What are the dimensions of x?" has an infinite number of answers, depending on one's personal preference. I have accordingly completely refrained from futile discussion of the latter question.

Dr. G. B. Brown. Dr. Yarnold stresses the fact that a constant a must always be included in the equation expressing an inverse-square law of force,

$$F = a\frac{q_1 q_2}{r^2},$$

and says that all we can say about the *dimensions* of a and q is that aq^2 has the dimensions of force multiplied by the square of a distance. Equally, then, according to this argument, all we can say about the *magnitudes* of a and q is that aq^2 will equal the product of the force measured and the square of the measured distance. Thus we never know the magnitude of an electric charge.

It does not seem generally realized that when we get over the second *impasse* by *defining the magnitude* of unit q in such a way as to make a unity, we also get over the first *impasse* by *defining the dimensions* of q in such a way that a is dimensionless. As I have said, "It can then, *and only then*, be dropped from the equation" (Brown, 1941, p. 424).

With regard to the derivation of the form of the inverse-square law in polarizable media: measuring the fall of potential across a condenser when a dielectric is introduced cannot tell us more than the *average* force which may be supposed to act on a unit charge as it is taken from one plate to the other. We certainly do not know the force on it at *any point*. If it is near a molecule we do not know what it is, and if it is not near a molecule, it must be in a cavity of some sort, and we want to know its shape. Then again, when the polarization of the medium is due entirely to two point charges we must remember that it will be differently polarized according as they are like or unlike. I do not think

there is any escape from some sort of cavity treatment such as I have attempted in an earlier paper (Brown, 1940).

Prof. R. O. Kapp. I think that in discussions on units and dimensions there is a tendency to confuse three different questions. The first of these concerns standards for purposes of copying, and is the only one to which experimental work has any relevance. The second involves metaphysical considerations. I doubt if it ever has any relevance. The third is concerned with labour-saving. I think this consideration alone ought to govern our choice of unit quantity in physics.

If we were untrammelled by the past we should seek to have only one unit for each physical quantity, and multiply this by sundry powers of ten, when it was inconveniently large or small. But custom and tradition are so strong that we cannot now dispense altogether with a multiplicity of units for certain quantities. We need, for instance, several units of weight and several units of length. The horse-power can hardly be suppressed in favour of the watt, or the calorie in favour of the erg. But our objection to such multiplicity, as well as our arguments in its favour, have nothing whatever to do either with experiment or metaphysics. The objection is that conversion factors are troublesome and lead to mistakes in arithmetic. The argument in favour is that it is less trouble to use conversion factors than to abandon established habits of thought.

This argument is weak when applied to existing systems of units in electricity and magnetism. Here we are saddled with three units for every quantity. We have the abvolt, statvolt, and volt. We have the abfarad, statfarad and farad, and so on throughout the whole of electricity and magnetism. Two of these three systems have been justified on the grounds that they are " natural ". The reason is that units in them are derived from the properties of a vacuum. But why should the properties of a vacuum be called more natural than those of any other medium? Is it merely because of the plentifulness of a vacuum in the Universe?

The real reason for our three units for every quantity has nothing to do with nature. It is because in electricity and magnetism, five unit quantities have been defined arbitrarily, and five is one too many. It is strange that this over-determination is defended by those who argue that the proper number of arbitrary determinations is three. The five which have been made in electricity and magnetism are the units of length, mass, time, permeability and permittivity.

This is the reason why the International Electro-technical Commission recommended unanimously in 1935 that the M.K.S. system should be adopted, in which only four arbitrary units are determined. These are the metre, the kilogram, the second and 10^7 times the permeability of a vacuum. In this system those magneto-electric units which are used most often, such as the ampere, volt, ohm, and farad, are identical with those in the practical system.

In conclusion I would like to point out that the fact that four arbitrary unit quantities have to be determined in electricity and magnetism is sufficient answer to those who argue that all physical quantities can be adequately expressed in terms of less than four fundamental dimensions.

Dr. C. Fox. A serious drawback to the theory of dimensions as practised at present is that it fails to distinguish between scalar and vector quantities, and I do not think that the theory can be clarified until modifications are introduced which make this distinction clear. Displacement is a vector quantity, and some vectors such as force, velocity, etc., show their vectorial nature in the theory of dimensions by having unity as their dimension in length. According to the formulae of Mr. Benham, the introduction of "Newton's constant" N gives force a dimension of -2 in length, which completely hides its vectorial nature. I consider this an objection to the theory proposed by Mr. Benham unless some other method of showing the vectorial nature of force is introduced in compensation.

Authors' replies:

Dr. G. D. Yarnold. There seems to be considerable confusion between the definition of a physical quantity and the statement of the way in which it is to be measured. Where a concept is so fundamental as to be incapable of definition, it is only possible to state the method to be adopted in its measurement. Thus, the concept *length* is indefinable, but length is to be measured by the ratio of two lengths, one of them a standard. However, this does not make length dimensionless. Dielectric constant, on the other hand, is defined as the ratio of two capacities, and is therefore defined to be dimensionless, but in this case the definition also indicates the method of measurement.

I entirely agree with Dr. Brown's remark about the magnitudes of a and q. But surely we are not at liberty to define the dimensions of q in such a way as to make a dimensionless? Electric charge appears to be something absolutely fundamental in the nature of things, whose dimensions cannot be fixed by reference to length and time. As regards Dr. Brown's second point, I can only reply that I have treated the medium from the classical point of view as continuous. At this stage there is no reason to enquire why a collection of atoms behaves as a medium.

I have read Dr. Guggenheim's comments and feel that his suggestions are thoroughly sound.

In reply to Professor Kapp, I agree that the definition of four, but not more than four, fundamental quantities is necessary in electromagnetism, but I regard the choice of the permeability of a vacuum in the M.K.S. system as regrettable.

Mr. W. E. Benham. I am interested in Dr. Fox's point of view, and agree that the theory of dimensions would be improved if it were possible thereby to distinguish between scalars and vectors. Dr. Fox fails to establish his objection

to the point of view of our paper, since he provides no criterion whereby the vectorial nature of a vector shall be judged. He starts by saying that the ordinary theory is deficient in the very respect for which he cites objection to the modification proposed. For his objection to hold it would be necessary that not merely *some* vectors . . . . have their dimensions in length unity, or, shall we say, odd, but *all* vectors. Conversely, scalars must presumably have *even* dimensions in length. Examples of vectors having even and scalars having odd length dimensions may readily be found. Thus a vector area $\boldsymbol{S}$ has dimensions L^2; similarly, regarding zero as an even number, the angular velocity ω, of dimensions 1/T, regarded as a vector drawn in right-handed manner along the axis of rotation, is of even dimensions in L; while energy per unit volume (a scalar, even when numerically equal to the vector pressure) has odd dimensions in L.

Now on our theory force has dimensions M/L^2. There would appear to be a case for including a unit vector $\boldsymbol{r}_1$, parallel to the direction of the force. More succinctly, if we regard the force as acting normally on a small area S,

$$[F] = \frac{M}{\boldsymbol{S}},$$

where $\boldsymbol{S}$ is the vector area. Some means of telling the direction of $\boldsymbol{F}$ would thus seem to be provided by regarding force as *mass per vector area*. This seems to be a real advance, for what could more aptly describe our *sensation* of force than this? I must express my acknowledgements to Dr. Fox for putting forward his point about vectors, without which the above dimensional expression for force might have been passed over.

In reply to Professor Dingle, it will be evident that we share Lanchester's views (1936) that "in physics the term dimensions denotes the kind or quality of a physical entity" and "that a single physical entity may possess more than one dimensional value is to the author unthinkable". While it is hardly within the scope of physics to discuss the nature of the fundamental units, physics can describe quantities not regarded as fundamental, such as temperature, in terms of the fundamental units. To this extent it is legitimate to talk of the nature of temperature. The fact that on the basis of thermometry temperature has dimensions of its own should make us suspect that thermometry does not go to the root of the matter. We then turn to the kinetic theory and to radiation, with the results indicated in the paper. It might, of course, be contended that one might choose to regard temperature as more fundamental than, say, either length or time. It would then be an error to talk of the nature of temperature, but one might perhaps talk of the nature of time if it were possible, on this basis, to reduce time to temperature measure. Prof. Dingle's suggestion for the measurement of temperature in interstellar space seems to be similar to the idea of using a sledge-hammer to knock in a tin-tack. It is possible to use the wrong tool. That a thermometer is suitable in laboratory experiments is due to the

fact that it is possible to bring the instrument into temperature equilibrium with its surroundings.

Regarding the problem of the variability of the earth's rotation, the standard of time need not be based, as Prof. Dingle suggests, on the distance travelled by light, despite the jeopardy into which the mean solar second is thrown by the supposed variation of the diurnal motion of the earth. For the temperature-controlled master-clocks used at observatories are capable of such precision that it is often the clock which supervises the astronomer rather than the reverse. One might therefore rest content, for example, with the definition of sidereal time as determined by star transit methods in any particular year (say 1942 or 2000 A.D.) and conserved in precision clocks until such time as the amazing refinements now attainable should perforce be superseded. Any slowing down of the earth's rotation (its amount is of the order of $\frac{1}{2}$ second per century) would then show up as a gradual divergence between astronomical and horological readings. Even if the ability of clocks to maintain regularity over long periods be in question, there are devices for overcoming this, and in view of compensations we should be adhering to a method of measurement which does not involve change of the time unit with length. Indeed, since length and time are separate fundamental units, measurement of time cannot involve length, though it may involve the ratio of two lengths or the coincidence of two marks. For example, the fact that we certainly use a length to measure temperature in a mercurial thermometer has no connection with the reciprocal length which, I am suggesting, connotes the intrinsic significance of temperature—the "thinness" of a pulse of electromagnetic radiation.

With regard to Professor Kapp's communication, as this is mainly directed to Dr. Yarnold, my comments will be but brief. From the point of view of the ordinary practising physicist, the question of labour-saving is of paramount importance, and choice of units should, indeed, be guided, if not governed, by this consideration. We are in favour of the M.K.S. system, and would also suggest rational units, making the so-called "permeability of a vacuum" equal to $(1{\cdot}256)10^{-6}$ henry per metre. There seems to us every reason for at length supporting Heaviside in his elimination of 4π from places where it ought never to have occurred.* The use of M.K.S. units, whether rational or irrational, should not be allowed, however, to influence us into thinking that the attribution of dimensions to μ and κ is anything but an artifice designed to enable us to forget about the presence of force constants.

REFERENCES

BROWN, G. B., 1940. *Proc. Phys. Soc.* **52**, 577; 1941. *Ibid.* **53**, 418.
GUGGENHEIM, E. A., 1941. *Nature, Lond.*, **148**, 751.
LANCHESTER, F. W., 1936. *The Theory of Dimensions and its Application for Engineers* (London), pp. vii and 125.
SOMMERFELD, A., 1935. *Z. tech. Phys.* **16**, 420.

* Compare Vigoureux, P., *Proc. Phys. Soc.* **52**, 611 (1940).

OBITUARY NOTICES

JULIUS RHEINBERG

RHEINBERG was a true amateur in the full sense of the word, for while engaged in ordinary commerce as his means of livelihood he devoted himself in all his spare time to the pursuit of science. His earlier work was with the microscope, and he published several papers in connection with the microscopic image. In particular, he devised a very fascinating way of exhibiting fine structure by means of a diaphragm placed in the lower focal plane of the condenser, of which the central area was one colour and the outer zone a contrasting colour. Thus the more oblique rays forming the image were of a different colour from the axial ones; whilst this probably has no advantage as regards the resolving power of the objective, it often makes the structure easier to observe. At the time when so much discussion took place upon the merits of the Abbe theory of the formation of the microscopic image, Rheinberg made a large number of experiments on diffraction in the endeavour to come to a positive conclusion on the points involved.

In the last war, he worked out a method of constructing fine graticules by photography which was of great value to the country—a method which the Government requested him to keep and work as a secret process. He once told me that he had obtained the fundamental idea from some of the early and forgotten methods of producing a photographic image—such as those, for example, described in Hunt's *Photography* (1854). These graticules were a great improvement on anything that had been made up to that time.

More recently he had turned his attention to stereoscopy. He made a number of experiments, but has published nothing of his results.

He was a regular supporter both of the Royal Microscopical Society and of the Physical Society, seldom missing either meetings, and was at various times on the Councils of both; he was twice a Vice-President of the former.

He had abounding energy and great enthusiasm for all new developments in Optics, which was his chief love.

Rheinberg himself was a most likeable man, who will be greatly missed by a large circle of friends. R. S. C.

MARK BENJAMIN

MARK BENJAMIN was killed in September 1941, when the aeroplane crashed in which he was returning from a visit to America. He was born in 1909, the eldest of a large working-class London family. He won a scholarship to the Regent Street Polytechnic, secured a high place in the L.C.C. open examination,

and started work at County Hall when he was 18. But his heart was set on scientific work; finding no scope for it in his position, he abandoned it, and in 1928 took a very subordinate post at the Research Laboratories of the General Electric Co., Ltd. Then he studied at evening classes and took a London degree in Honours Physics. His abilities were soon recognized. He was attached to the Valve group, and worked first with A. L. Reimann and then with C. J. Smithells. When both of these had left, in 1938, he was put in charge of one section and became responsible for the co-operation of the whole group with the factory.

To great intellectual power, Benjamin added tireless energy and an irresistibly infectious enthusiasm. Tact he would not have claimed; but his criticism was always so constructive, and so evidently inspired only by a desire to achieve the common end, that it was tolerated, and even welcomed, almost everywhere. He was the ideal industrial scientist, combining the abstract intuition of the investigator with the practical sense of the manufacturer.

In his short career Benjamin collaborated in several papers of general interest. With H. P. Rooksby he studied the thermionic emission from mixed barium and strontium oxides; they reached the conclusion that the function of the strontia is to form a solid solution with the baryta to which the monatomic layer of barium adheres more firmly than it does to pure baryta. This work won him the degree of Ph.D. With R. O. Jenkins he investigated the diffusion of barium on metal surfaces at room temperature; but here their conclusions were shown by Becker to be too general. His latest work, also with Jenkins, concerned the autoelectronic discharge from fine metal points, which can act as a powerful electronic microscope, revealing the minute structure of the surface from which it takes place. Their beautiful experiments, some still in the process of publication, greatly extended the work of Müller; in particular they produced evidence of the surface migrations of metal atoms under high electric fields at temperatures far below the melting point (e.g. tungsten at 1200° K.).

Benjamin leaves a widow, to whom his work owes much, and two young children.

N. R. C.

TOM HARRIS

T. Harris, who died on 15 January 1942, at the comparatively early age of 56, was born in London on 4 November 1885. He was educated at Upper Latymer School and later at the Royal College of Science, where he took a degree in Physics. From London he went to Cambridge as an advanced student and, although strictly a London man, it was always Cambridge that was his intellectual home. I could never be quite sure whether his attraction was in the scholarly atmosphere of the picturesque colleges, or in the inspiration that he derived from that great physicist, to whom so many owe so much, Sir J. J. Thomson.

On leaving Cambridge in 1911 Harris was appointed lecturer on the Physics Staff of Queen Mary College, London, but it was not very long before the war of 1914 claimed him for a commission in the Royal Engineers. He saw service in England and France—mostly in connection with meteorological work.

At the end of the war, Harris returned to Queen Mary College. He was an excellent colleague, and the years that immediately succeeded were, to me, and, I think, to him, of the pleasantest. It was in the discussion of physical problems that arose as a matter of general interest or out of research that Harris showed up so well. I always admired the way in which he would never rest content until he was quite clear about the problem under discussion. His mind was not of the type that would hurry a piece of research so as to be first in the field. He cared little for notice. To be clear himself and to explain clearly to others seemed to be his only object.

Harris, like myself, left Queen Mary College in January 1921, and joined the Research Dept. at Woolwich. Although we were now in widely separated branches, I used to see him frequently. It was here, of course, that the major part of his research work was done. It is a pity that its confidential nature makes discussion of it impossible, for Harris had a power of attacking a problem both analytically and experimentally, and I can remember the enthusiasm with which his assistants spoke of his powers.

Harris's work in the present war took him once again to Cambridge. Knowing his pleasant memories, I could not imagine more congenial surroundings, but, alas! I knew little about the trying nature of his work and of his failing powers.

His loss will be deeply felt by all who knew him.

He leaves a widow and a grown-up son and daughter.

G. D. W.

REVIEWS OF BOOKS

University Physics. Part III, *Light*, by F. C. CHAMPION. Pp. 172. (London and Glasgow: Blackie and Son, Ltd., 1941.) 5*s.* 6*d.* net.

The value of a text-book on Light can be very largely made or marred by the diagrams it contains. Optics is a notoriously dry subject for most students, but well-drawn and accurate diagrams can go a long way to lighten the algebra and make it intelligible. Unfortunately, the illustrations in this book have been very carelessly drawn. There are at least 24 diagrams that are incorrect in some way or another—rays refracted in impossible directions, incident and reflected rays on the same side of the normal, eyes placed at the wrong position to see the image, astigmatic images not shown at the correct intersections, object and image points that could not possibly be conjugate to one another, innumerable interference fringes where only a few actually can be formed, and so on. All this makes the task of the student very difficult. It is not assisted by the hesitancy with which the author uses his sign convention, leading on p. 41 to the curvature term in the equation for the focal length being derived as $(1/r_1 + 1/r_2)$, while on p. 28 it had been derived as $(1/r_1 - 1/r_2)$—a discrepancy not to be attributed to the printer. The book as a whole is altogether too sketchy to be of real value, the subject matter which the author attempts to cover being far too wide for the 150 pages devoted to it. If there was a need for a new book on Light, it certainly has not been met in this volume.

W. D. W.

The Twin Marchant Calculating Machine and its application to survey problems. By L. J. COMRIE, Ph.D., M.A. Pp. 40. (London: Scientific Computing Service, Ltd., 1942.) 10*s.*

This large booklet is actually the third, but a largely rewritten and re-titled, edition of a work which appeared in 1938. The description applies not only to a twin but also to the single *Marchant* machine. The twin machine has been cleverly constructed, apparently with complete success, by connecting two singles back to front; the two main operating handles are replaced by a single crank acting through gearing on the two multiplying mechanisms. At the moment this is apparently the only twin machine available, though it is expected that the Swedish *Facit* machine will be obtainable after the war. It is noteworthy that the greater number of modern machines in use for general calculation are developments of the original design by Odhner, a Swede who lived in St. Petersburg; the *Leibniz bar*, in spite of its silent working, has evidently not been found capable of similar development.

With this and many other machines and a staff of highly trained assistants, Dr. Comrie is now in a position to undertake the computation of the most complex functions and the resolution of onerous problems, such as those of least squares with more than, say, 20 unknowns. These problems could not be described in a booklet of reasonable dimensions; accordingly no more than simple operations are considered. Particularly valuable are the hints given as to the proper use of the machines and the short cuts available. The twin machine is, of course, used especially for the common case of two or more multiplicands with a common multiplier.

J. T. McG.

Meteorological Instruments, by W. E. KNOWLES MIDDLETON. Pp. viii + 213. (Toronto: The University of Toronto Press; London: Sir Humphrey Milford, at the Oxford University Press, 1941.) 11*s.* 6*d.*

No comprehensive English text on meteorological instruments has been published in the present century; we have had only articles in Glazebrook's *Dictionary* and the elementary though well-conceived *Meteorological Observer's Handbook*, now in a thirteenth edition. Kleinschmidt's *Handbuch der Meteorologische Instrumente* (1935) is the only thorough treatment available, and this will hardly lose its place with the appearance of the book under review, though the latter, much smaller in bulk, supplements Kleinschmidt in several practically important respects and has the merit of being more critical and less diffuse. Middleton's work is directed essentially to the meteorologist responsible for day-by-day observation of the common meteorological quantities, those which go to making the surface and upper-air synoptic weather charts. Thus he deals with the measurement of barometric pressure, air temperature, humidity, precipitation and evaporation, wind near the earth's surface and in the free atmosphere, cloud characteristics, sunshine duration and radio exploration of the free atmosphere. Less general but fundamental measurements on solar and terrestrial radiation, atmospheric opacity, etc., are not considered, but a short section is devoted to the measurement of cloud-particle size and water content.

There are excellent features to the book. Thus the author states clearly the principles of good meteorological instrument design, principles which many investigators learn only by long experience in field measurement. He makes a point of the accuracy of observation which is worth attaining, and that which any instrument does in fact give—as opposed to figures based on unconsciously idealized conditions. The treatment of a property of a class or sub-class of instruments is given point to by a few numerical figures; these help to clinch an argument, on instrument lag for example, and add much to the usefulness of the book. And one admires the balanced treatment of radio-sondes, which have been developed so rapidly in recent years that the balanced view is difficult to obtain. Also, as a Canadian meteorologist in close touch with British and American practice, the author claims to present the better features of both.

One's criticism of the book is that, within its professed limits, the treatment is not sufficiently thorough and up to date, and an addition to the two hundred pages should have been made to fill the lacunae. Hygrometry is an example. Meteorologists, indeed physicists generally, should be made aware of Simon's and of Powell's work showing how the reduction of wet-bulb size may almost eliminate the effect of ventilation on the psychrometric "constant". The hair hygrometer is well treated, but only passing reference is made to the superior qualities of gold-beater's skin, and for neither instrument is it explicitly stated that at sub-freezing temperatures the indications are of relative humidity with respect to super-cooled water and not to ice.

Other omissions relate to aneroid barometers—the desirability of using beryllium bronze to avoid elastic errors in aircraft and radio-sonde instruments; to radiation effects on meteorological instruments—no general treatment is attempted; and to the effect of Reynolds number on the drag coefficient of balloons.

The book shows evidence that the subject, from a period of relative stagnation, is once again very live. A meteorologist may now have before him in his working room indications or records of what is happening "out there" in the atmosphere, at least for the more important variables. Other notable advances are in balloon theodolite

design, cup and pressure-plate anemometry, a recording dew-point hygrometer and the radio-sonde already noticed.

The text is concisely written, well indexed and referenced and attractively presented. There are few serious errors or misprints, but two may be noted. The error in temperature indication due to airspeed is given (p. 54) proportional to air density, while it is, in fact, independent of the latter ; and there is confusion in nomenclature in the treatment of balloon lift (p. 149). The experimental meteorologist will be glad to possess the book, though wishing for more, and the general physicist will find it of use when measurements of the qualities of an atmosphere are to be made. P. A. S.

Reports on Progress in Physics, Volume VIII (1941). General Editor, W. B. MANN. Pp. iv+372. (London : The Physical Society, 1942.) 25*s*.

This new volume contains articles on : *The teaching of theoretical physics in universities* (10 pages), by M. Born ; *The theory of the specific heat of solids* (20 pages), by M. Blackman ; *High-speed centrifuging* (18 pages), by J. W. Beams ; *Applications of spectroscopy to combustion* (21 pages), by A. G. Gaydon ; *Friction between solid bodies* (19 pages), by R. Schnurmann ; *The general physical constants* (45 pages), by R. T. Birge ; *Recent television developments* (51 pages), by V. K. Zworykin and R. E. Shelby ; *Instrumental technique in astrophysics* (14 pages), by A. Hunter ; *Dispersion in the far infra-red* (12 pages), by L. Kellner ; *The M.I.T. wavelength project* (19 pages), by G. R. Harrison ; *Molecular electronic spectra, dispersion and polarization* (43 pages), by R. S. Mulliken and C. A. Rieke ; *Symmetry properties of nuclear levels* (44 pages), by E. P. Wigner and E. Feenberg ; *Photometry* (20 pages), by H. Buckley ; *Discharge phenomena in gases* (30 pages), by R. W. Lunt, A. von Engel and J. M. Meek. An author index (5 pages) is provided.

With Volume VIII is issued the List of References for the *Report* by F. EIRICH on "Viscosity of Substances of High Molecular Weight in Solution" which appeared in Volume VII (1940).

RECENT REPORTS AND CATALOGUES

Rotameters. Pp. 5. ROTAMETER MANUFACTURING CO., LTD., Vale Road, Portslade, Sussex.

Kryptoscreen X-Ray Paper and Brytex Intensifying Leaf. Pp. 6. ILFORD, LTD., Ilford, London.

The Cambridge Bulletin (No. 88, Spring 1942). Pp. iii+16. CAMBRIDGE UNIVERSITY PRESS, Bentley House, 200 Euston Road, London, N.W. 1.